Airbnb, Hotels, and Localized Competition*

Maximilian Schaefer
Yale University
maximilian.schaefer@yale.edu

Kevin Ducbao Tran
University of Bristol
kevin.tran@bristol.ac.uk

February 14, 2022

Abstract

The rise of online platforms has disrupted numerous traditional industries. A prime example is the short-term accommodation platform Airbnb and how it affects the hotel industry. On the one hand, consumers can profit from Airbnb due to an increased number of choices and lower prices. On the other hand, critics of the platform argue that it allows professional hosts to operate de facto hotels while being subject to much laxer regulation. Understanding the nature of competition between Airbnb and hotels as well as quantifying consumer welfare gains from Airbnb is important to inform the debate on necessary platform regulation. In this paper, we analyze competition between hotels and Airbnb listings as well as the effect of Airbnb on consumer welfare. For this purpose, we use granular daily-level data from Paris for the year 2017. We estimate a nested logit model of demand that allows for consumer segmentation along accommodation types and the different districts within the city. We extend prior research by accounting for the localized nature of competition within districts of the city. Our results suggest that demand is segmented by district as well as accommodation type. Based on the parameter estimates, we calibrate a supply-side model to assess how Airbnb affects hotel revenues and consumer welfare. Our simulations imply that Airbnb increases average consumer surplus by 4.3 million euro per night and reduces average hotel revenues by 1.8 million euro. Furthermore, we find that 28 percent of Airbnb travelers would choose hotels if Airbnb did not exist.

Keywords: hotel industry, short-term rentals, localized competition, consumer welfare, sharing economy, peer-to-peer markets, Airbnb

JEL codes: D4, D6, L1, Z38

*We thank Matthias Hunold, Reinhold Kesler, and Ulrich Laitenberger for providing us with access to hotel price data. We thank Tomaso Duso and Hannes Ullrich for helpful comments. We further appreciate helpful comments by seminar participants at DIW Berlin and Católica Lisbon. All remaining errors are our own.

1 Introduction

In fall 2007, the co-founders of the peer-to-peer short-term accommodation platform Airbnb hosted their first guest in their San Francisco apartment.[1] As of 2016, the platform listed three times as many rooms as the largest global hotel chain, Marriot International (Haywood et al., 2017). This rapid rise was accompanied by controversial policy debates. On the one hand, consumers most likely benefit from Airbnb due to the increased variety in choices and the competitive pressure the online platform puts on traditional hotels, resulting in lower prices. On the other hand, the hotel industry argues that many of the hosts on Airbnb effectively run small hotels without facing the same level of regulation.[2] In this discussion, it is important to understand what is the actual nature of competition between Airbnb listings and hotels and to quantify how consumers profit from the presence of the peer-to-peer accommodation platform.

In this paper, we combine data on hotel and Airbnb demand in Paris in 2017 to estimate a discrete choice model of accommodation demand. Based on the estimates, we assess the nature of competition between Airbnb and hotels. In simulations, we next quantify how Airbnb affects consumer welfare as well as hotel revenues and demand.

One main contribution of our paper is that we can account for variation in demand within the city. Our data contains daily-level information on hotel and Airbnb demand in Paris. Furthermore, because we have information on the location of hotels and Airbnb listings within the city, we can allow for segmentation of demand not only along the type of accommodation but also its location. Consequently, we can assess how localized competition is within the city and include this localization in our counterfactual simulations accordingly. Our data suggest that local, short-term demand variation within the city plays an important role in the accommodation market. Prior research only uses variation across cities and time

[1] See https://news.airbnb.com/fast-facts/ (accessed: July 17, 2020).

[2] A 2017 study commissioned by the American Hotel & Lodging Association (AHLA) highlights the increasing importance of multi-unit hosts on Airbnb. This prompted AHLA President Katherine Lugar to call on Airbnb to “crack down on the illegal hotels that it facilitates.” See https://www.ahla.com/press-release/new-study-shatters-airbnb-homesharing-myth (accessed: July 3, 2020).

to identify their estimates (Zervas et al., 2017; Farronato and Fradkin, 2018).

In 2017, Paris was the second most popular Airbnb destination in Europe, not far behind London.[3] Therefore, our analysis for Paris contributes important pieces of evidence for the European debate on the regulation of short-term accommodations. Existing research mostly focuses on the US.

Our results suggest that consumers perceive Airbnb listings and hotels as substitutes, but substitution is stronger within each of the two types of accommodation. In particular, consumers value hotels more highly than Airbnb listings, on average. Our findings also show that competition is localized. Consumers' preferences are correlated for accommodations in the same district. Consequently, accommodations in the same district are closer substitutes than accommodations in different districts. For example, we find that a ten percent price increase of five-star hotels leads to a one percent increase in demand for Airbnb listings in the same district, while the demand for accommodations in other districts increases by only 0.1 percent.

Based on the estimated demand parameters, we quantify the gain in consumer surplus from Airbnb in counterfactual simulations. Our results suggest an average gain in consumer surplus of roughly 4.3 million euro per night (or 31 euro per traveler). This gain in consumer surplus is a result of both the increased choice from the availability of Airbnb as well as of hotels charging lower prices with increased competition from Airbnb. Ignoring this price effect, we find that the average consumer surplus gain from Airbnb amounts to approximately 3.1 million euro per night. These gains in consumer surplus are heterogeneous over time: When hotels are near their capacity limits, consumer surplus gains from Airbnb total up to eight million euro per night (including the gain from lower hotel prices). Our results suggest that Airbnb is most beneficial to consumers during periods of high demand when hotels would charge higher prices absent Airbnb.

For our analysis, we combine data from three main sources. For Airbnb demand and

[3]See https://www.statista.com/statistics/957312/airbnb-leading-european-destinations/ (accessed: July 17, 2020).

prices, we use web scraped data obtained through AirDNA.[4] For hotel demand, we use a monthly survey conducted by the French National Institute of Statistics and Economic Studies (INSEE)[5] that asks a sample of French hotels for their daily occupancy rates. Because these data do not include hotel prices, we combine them with web scraped hotel prices from Booking.com, a hotel booking platform.[6] By combining these different data sets, we can estimate a joint model of hotel and Airbnb demand.

For the estimation, we define products based on their type (hotel or Airbnb listing), their quality, and the district in which they are located. In our main analysis, we propose a two-level nested logit model in which the first nesting tier represents the choice between different districts in the city and the second nesting tier represents the choice between hotels and Airbnb listings within a district. This nesting structure allows for correlated preferences by geography (districts) and accommodation type while remaining tractable.

Using the parameter estimates, we next simulate consumer welfare, hotel prices, and hotel demand in counterfactual scenarios. The main counterfactual we are interested in is a world in which Airbnb is absent. To simulate this counterfactual, arguably the biggest hurdle is simulating hotel prices. The main problem is that hotels' profit maximization most likely needs to account for the capacities of nearby hotels as well as their own capacity constraints. Analytically solving a price equilibrium in this setting would be complex. Therefore, we propose a heuristic approach, to solve for the equilibrium numerically. The basic idea is to iteratively assign predicted demand to hotels up to their capacity constraints and repredict the remaining excess demand. At the same time, hotel prices that are consistent with the assigned demand need to be found.

Our work is most closely related to the work by Zervas et al. (2017) and Farronato and Fradkin (2018). Using data from Texas, Zervas et al. (2017) find that an overall increase in Airbnb supply reduces hotel revenues by approximately ten percentage points in their sample,

[4] See `https://www.airdna.co/` (accessed: July 3, 2020).

[5] See `https://insee.fr/en/accueil` (accessed: July 3, 2020).

[6] The data collection was initially done as part of the work on Hunold et al. (2020). We are grateful to the authors for providing us with access to the data.

with lower-priced and non-business hotels being disproportionately strongly affected. They show that the revenue impact is mainly driven by a reduction in prices as opposed to a reduction in the number of rooms booked.[7]

Closest to our work, Farronato and Fradkin (2018) employ a random coefficient discrete choice model and estimate the demand and supply parameters using data for several US cities. Consistent with our results, they find that consumers profit most from Airbnb in locations and at times when hotels are capacity constrained. Because they use data on the city level, they identify capacity constraints when demand for the entire city increases (e.g. for New Year's Eve in New York City). In our granular data, we show that capacity constraints in individual districts in the city are actually binding a lot more frequently than those for the entire city. Thus, combined with the localized level of competition that our results suggest, accounting for variation within a city seems important to analyze the accommodation market.

Furthermore, our research is related to a growing body of literature studying competition between peer-to-peer platforms and incumbent firms. Seamans and Zhu (2014) and Kroft and Pope (2014) analyze the impact of Craigslist on incumbent industries. Aguiar and Waldfogel (2018) assert that streaming affects music sales using Spotify data. Cohen et al. (2016) and Lam et al. (2020) study consumer surplus from Uber, while Hall et al. (2018) address the question of Uber's complementary to public transport systems. Cramer and Krueger (2016) assess the relative efficiency of Uber drivers compared to taxi drivers.

More broadly, our paper is also related to the literature studying externalities related to the platform economy. One research strand studies how Airbnb affects housing markets (Horn and Merante, 2017; Koster et al., 2018; Garcia-López et al., 2019; Barron et al., 2020; Duso et al., 2020; Valentin, 2020). Basuroy et al. (2020) analyze Airbnb's impact on local neighborhoods. Our paper is also broadly related to the emerging literature on online platforms and surge pricing (Cachon et al., 2017; Castillo, 2019; Guda and Subramanian,

[7]In a similar exercise, Neeser (2015) uses data from three Nordic European countries. Unlike Zervas et al. (2017), the author does not find a statistically significant impact of Airbnb on hotel performance.

2019) as we show that Airbnb is particularly valuable for consumers when hotels charge higher prices.

The remainder of the article is organized as follows. In Section 2, we describe some institutional details of the Parisian accommodation market and introduce the various data sources that we use in more detail. In Section 3, we present our empirical strategy by showing selective descriptives, introducing our model, providing some estimation details, and finally discussing identification of the key model parameters. In Section 4, we report the estimated parameters as well as demand elasticities. In Section 5, we outline our simulation procedure and present the results of the counterfactual analysis. Section 6 concludes.

2 Institutional Details and Data

In our analysis, we assess the competition between Airbnb and hotels using data for the city of Paris in 2017. Therefore, in this section, we first briefly present some facts about the accommodation industry in Paris. Subsequently, we introduce our main data sets in some detail. Finally, we describe how we define products for our analysis.

2.1 The Short-Term Accommodation Market in Paris

According to Mastercard's Global Destination Cities Index, Paris is the city with the third largest number of international visitors worldwide in 2017.[8] Only Bangkok and London attracted more international travelers. In the 2019 edition of the same report, Paris even surpassed London and was the second most visited city in the world.[9] Accordingly, Paris is also the second most popular Airbnb destination in Europe, not far behind London, in

[8]See https://mastercardcontentexchange.com/newsroom/press-releases/2018/big-cities-big-business-bangkok-london-and-paris-lead-the-way-in-mastercards-2018-global-destination-cities-index/ (accessed: July 17, 2020).

[9]See https://newsroom.mastercard.com/wp-content/uploads/2019/09/GDCI-Global-Report-FINAL-1.pdf (accessed: July 17, 2020).

2017.[10] Therefore, studying the competition between Airbnb and hotels in Paris amounts to studying it in one of the most important short-term accommodation markets not only in Europe, but the world.

Paris has introduced regulations that limit the number of days an apartment can be rented out during a calendar year. Since October 2017, the City of Paris requires hosts of entire apartments on Airbnb to obtain and display a registration number. Since January 2020, hosts of entire homes can no longer rent out their apartments for more than 120 days in a year.

2.2 Data

Our goal is to estimate a joint model of Airbnb and hotel demand in Paris. Therefore, the main necessary components for our analysis are demand and prices for hotels and Airbnb listings. To obtain this information, we combine three different data sets that we describe in the following.

Hotel Demand Data To assess hotel demand, we use data provided by National Institute of Statistics and Economic Studies (INSEE). Each month, INSEE surveys a random sample of French hotels and asks for information regarding occupancy and origin of guests.[11] Our main question of interest concerns the number of booked rooms for each day of the survey month. Based on this self-reported daily occupancy, we create our measure of hotel demand. For each hotel, the data also include the star rating, the number of available rooms, and the district.[12] Based on the hotel responses, INSEE imputes the occupancy for non-respondents as well as non-sampled hotels. Consequently, the data contain (imputed) quantity information for the

[10] See https://www.statista.com/statistics/957312/airbnb-leading-european-destinations/ (accessed: July 17, 2020).

[11] The population of hotels is first stratified by region, star rating, and type (chain or independent). Within each stratum, random samples are drawn. INSEE reports that a sample of approximately 12,000 out of 18,000 hotels is contacted in the national sample. See https://insee.fr/en/metadonnees/source/operation/s1480/processus-statistique (accessed: July 16, 2020).

[12] The exact address of the hotels is also available but since the data are classified as sensitive, we are prohibited from geocoding the addresses.

entire population of hotels.

Hotel Price Data The hotel quantity data from INSEE lacks an important component for demand estimation: hotel prices. Thus, we use web scraped hotel price data from the hotel booking platform Booking.com. For each night in 2017, the data include the prices for each hotel available on Booking.com two weeks, one week, and one night prior to the night of interest. The prices reported are those shown as a result of a search for a one-night stay of two persons. If there are multiple room categories, the reported price is that for the cheapest available category.[13] To match this price information to the hotel quantity data, we use hotel names. This match is possible for 983 of the 1,597 hotels in the quantity data.[14] Due to the occasional instability of the web scraper, the number of observed hotels varies. Therefore, we exclude dates from the analysis for which we do not observe prices for at least 65 percent of the total number of rooms. As a result, 232 dates remain in our analysis.

Airbnb Data For our analysis, we require data not only on the supply of Airbnb listings in Paris, but also on the corresponding demand. Obtaining a measure of Airbnb supply is possible by web scraping the Airbnb website. More specifically, by searching the website for listings in Paris without restricting the search to any specific dates, all listings that are available for booking at some date in the future can be found. However, obtaining a measure for Airbnb demand is more difficult. The reason is that Airbnb does not display whether a listing was booked for a given date. In principle, for each listing a calendar is shown that indicates whether it is available for booking on a given date or not. However, if a listing shows as unavailable, this can mean that it is booked for that date or that the host blocked it for private reasons.

Therefore, we use web scraped Airbnb data available from AirDNA.[15] The data include all listings that were available for booking in Paris in 2017. The advantage of the AirDNA

[13] See Hunold et al. (2020) for more details on the data gathering procedure.
[14] We compare matched and non-matched hotels in Appendix A.1.
[15] See `https://www.airdna.co/` (accessed: July 3, 2020).

data is that they include information on whether a listing was available, blocked, or booked on each date. This information is imputed by AirDNA. They are able to do so because they gathered data from Airbnb when the website itself still displayed whether an unavailable listing was booked or blocked. They use this information to predict whether an unavailable listing in the present data is most likely booked or blocked. While imperfect, this measure for Airbnb demand is the best available that we are aware of with the exception of data obtained from Airbnb directly. In addition to the imputed booking status, the data include the price, the location, and further listing-specific characteristics that are observable on the Airbnb website.

For the analysis, we also need a measure of listing quality to account for the impact of quality on demand. While hotels have a natural quality measure (their star rating), this is not the case for Airbnb listings. Therefore, to define quality categories for Airbnb accommodations, we follow Farronato and Fradkin (2018). We regress Airbnb prices on time and listing fixed effects. Subsequently, we divide the estimated listing fixed effects by the quartiles of their distribution and categorize each Airbnb listing into one of the four groups according to their estimated fixed effect. As a result, we create four Airbnb quality categories.

2.3 Product Definition

In principle, we observe individual transactions for Airbnb listings. For hotel demand, we have the occupancy of each hotel on each date. Therefore, it is conceivable to conduct our demand estimation using a disaggregated discrete choice model. A major advantage of such a disaggregated analysis would be that it would allow us to take into account the specific location of each product. However, while we do have the precise location for each Airbnb listing, we do not observe this for hotels due to the sensitivity of the data. Instead, we only know in which district each hotel is located.[16] Therefore, if we were to estimate demand on

[16] In principle, it would be possible to geocode each hotel in the hotel price data based on its name. This would yield information on the location of each hotel for which we have prices. While this geocoding

a disaggregated level, we would nevertheless only be able to use geographic information on the district level.

Another potential advantage of disaggregated estimation is that variation in consumers' choice sets can help to empirically identify parameter estimates. However, we do not observe the time at which each consumer booked their accommodation. Therefore, we do not observe if consumer choice sets have changed due to accommodations having been booked. Thus, including variation in choice sets in the model would require additional, unverifiable assumptions.

Finally, one disadvantage of disaggregated estimation is its immense computational burden. In particular, with potentially large choice sets that often occur in the context of online platforms, disaggregated discrete choice estimation can quickly become computationally infeasible.

Due to these concerns, we decided to aggregate the data for our analysis. For the aggregation, we define a product as a combination of type (Airbnb listing or hotel), quality category, and district. As a result, we circumvent the problems of disaggregated analysis discussed above while still being able to consider geographic heterogeneity within the city (on the district level).

As quality categories, we use the star ratings for hotels and the fixed effects categorization due to Farronato and Fradkin (2018) and described in Section 2.2 for Airbnb listings. As prices, we use the average transaction price for each product for a given date of stay. Note that these prices are missing for a fraction of hotels. As a result, the average hotel prices are based on only those hotels for which price data is available. However, when calculating the aggregate demand for each product, we use the entire sample of hotels. If we used only those hotels for which we have prices, the demand for hotels with no price data would wrongly be

based on names would probably not work perfectly throughout, we expect that it should work well for most hotels. However, since INSEE does not permit geocoding of the hotel addresses, it is unethical to do so in a workaround and would likely violate the INSEE use agreement. Furthermore, this approach would mean that we can only use 983 out of 1,597 hotels (those for which we can successfully match the price to the quantity data).

assigned to the outside good. Therefore, an implicit assumption in our aggregation is that the average prices of hotels with and without price information are comparable.

With this aggregation procedure, our analysis is based on a maximum of 180 possible products per market.[17] The product market shares are defined as the ratio of the number of booked rooms over the market size for a particular night. We treat each night as a separate market. This assumption is necessary, because, as discussed earlier, the hotel quantity data do not allow us to distinguish between individual bookings.

3 Empirical Strategy

In this section, we discuss our empirical strategy. We begin by presenting selected descriptives that highlight particularities of the Parisian short-term accommodation market that are important for identification and estimation. Next, we describe the model on which the estimation is based. Subsequently, we set forth some estimation details. Finally, we discuss some issues concerning the identification of the model parameters.

3.1 Descriptives

For a better understanding of the data and to highlight issues of importance for estimation and identification, we first present some descriptives. In particular, we show how Airbnb and hotel supply changes over time. Further, we document evidence for the importance of localized competition in the short-term accommodation market in Paris. The results presented in this subsection motivate our choice of the demand model and the identification strategy that we discuss in subsequent subsections.[18]

The main components to any demand estimation are supply, demand, and prices. Therefore, we begin our descriptive analysis by examining the supply side of the short-term accom-

[17] These 180 products result as a combination of 20 districts and nine type-quality combinations (four Airbnb quality categories and five possible hotel stars).

[18] We present summary statistics by hotel and Airbnb quality category in Appendix A.1.

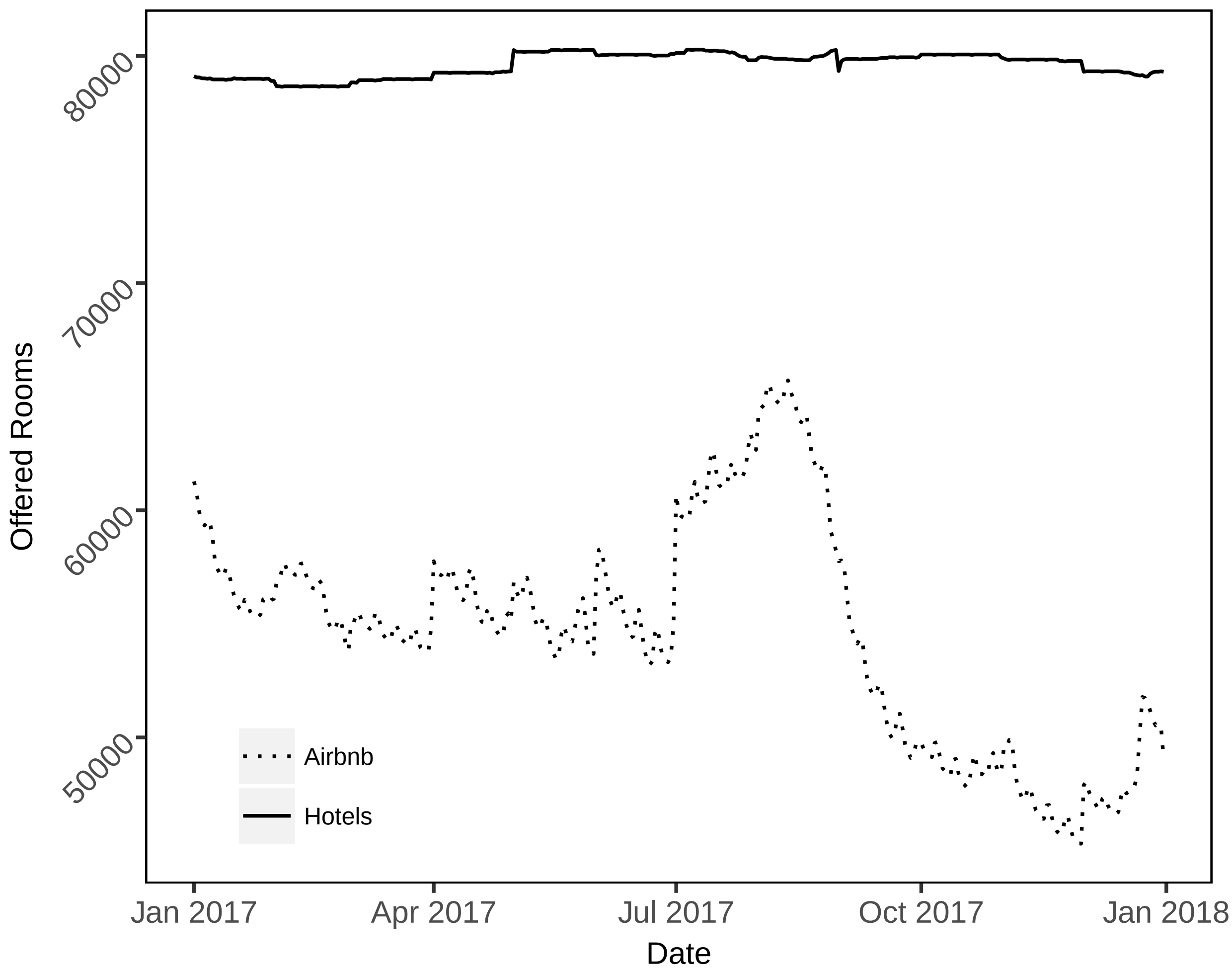


Figure 1: Offered rooms by date, split by hotels and Airbnb listings

modation market in Paris. Figure 1 presents the number of rooms available in Paris for each date in 2017, split by hotels and Airbnb listings. Airbnb listings make up approximately 40 percent of the overall room capacity in Paris, on average. However, while hotel room supply is almost constant over time, the supply of Airbnb listings fluctuates both in the short- and long-runs. Around October 2017, the number of Airbnb listings in the city decreases substantially. This decrease is likely due to the introduction of a mandatory registration number display for Airbnb listings in Paris starting in October 2017.[19]

Turning to demand over time, Figure 2 presents the number of rooms occupied for both

[19]This decrease in Airbnb listings is arguably exogenous to demand for accommodations in Paris. However, focusing exclusively on the period around the policy change is not feasible for our analysis, because most of the observed transactions take place before October 2017.

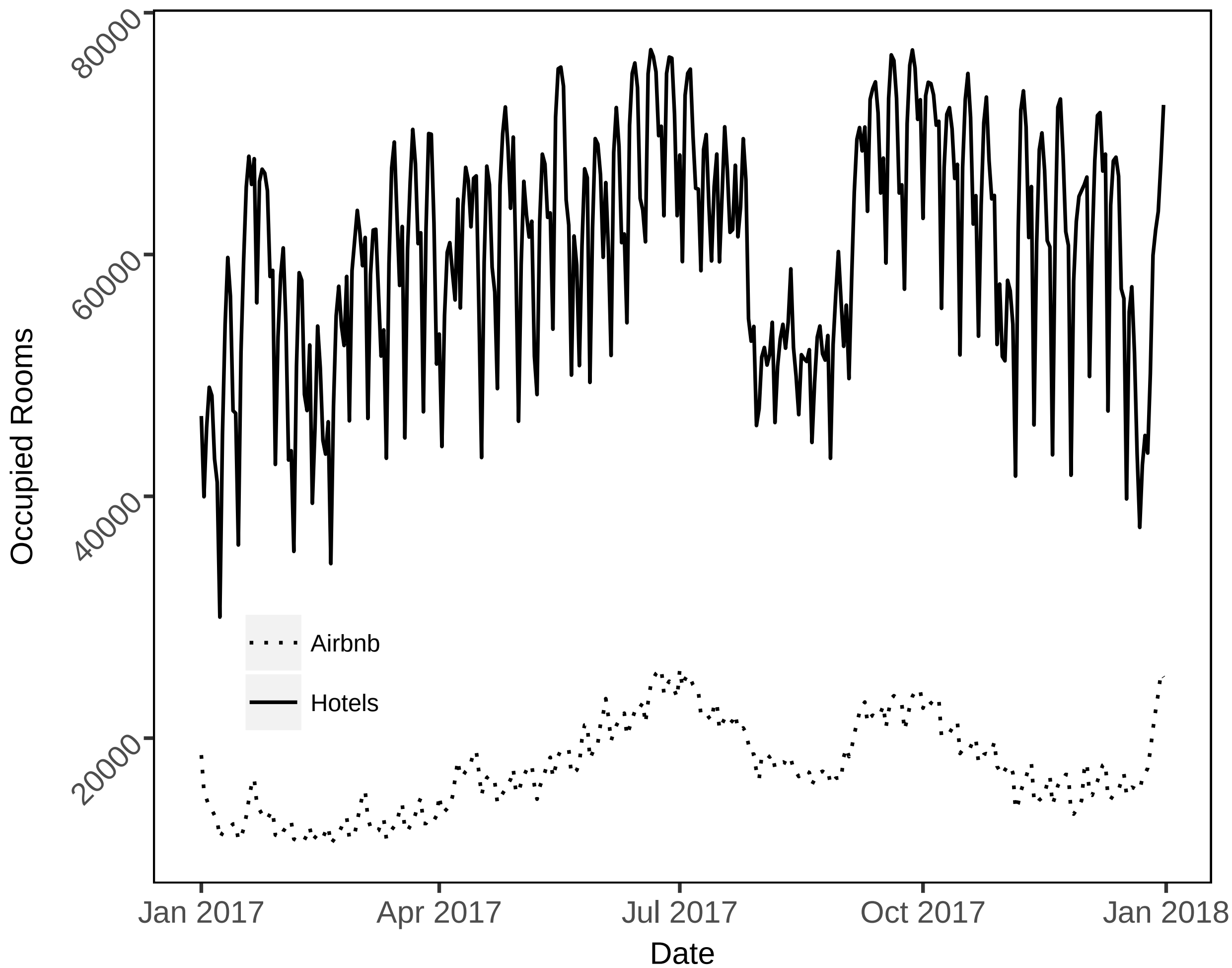


Figure 2: Occupied rooms by date, split by hotels and Airbnb listings

types of accommodation. Contrasting the supply patterns, demand for hotels varies more strongly. Usually, hotel demand is higher during weekdays than weekends. Airbnb demand, on the other hand, is more stable in the short run. Despite this difference in variance, demand for Airbnb and hotels appear to move roughly in sync. This joint movement suggests that demand for the two types of accommodation is, at least, not completely independent from one another.

Next, we consider descriptives that illustrate the relationship between Airbnb and hotel prices. Figure 3 shows the average daily prices for the two types of accommodation.[20] The average hotel price exhibits larger variation and follows the demand patterns shown in

[20] Note that, for hotels, we only include the days for which we have sufficient price information.

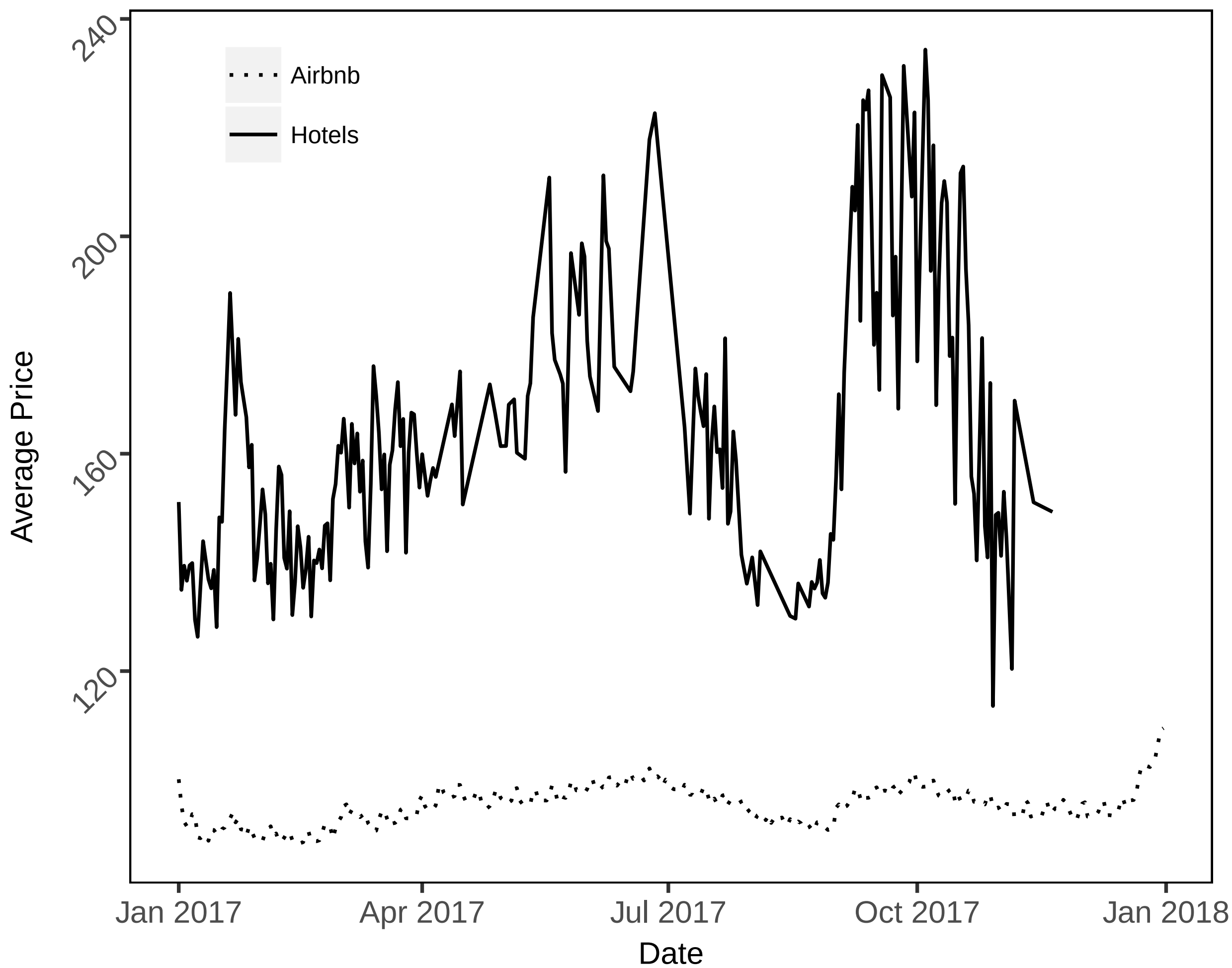


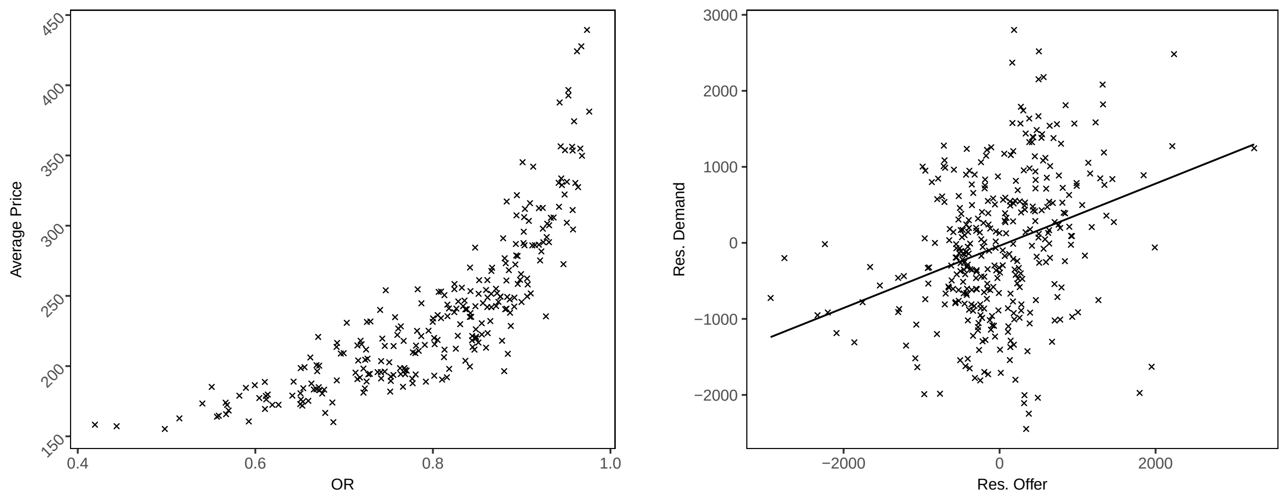
Figure 3: Average hotel and Airbnb prices over time

Figure 2. In comparison, the variation observed for Airbnb prices is negligible. This figure suggests that there are fundamental differences in price-setting behavior of hotels and Airbnb listings. This insight is important, especially when we simulate hotel and Airbnb prices in counterfactual scenarios in later sections.

The next set of descriptives combines an analysis of demand and price setting. Figure 4a illustrates how hotel prices correlate with the average occupancy ratio (i.e. the number of booked rooms over the total number of available rooms). For exposition, this figure shows the relationship for four-star hotels in the first district. However, this pattern can be observed more generally for other types of accommodation and other districts as well. In general, hotel prices increase with higher occupancy ratios. However, the relationship appears to be non-linear. Specifically, when hotels reach their capacity limits, prices increase more sharply. When simulating hotel price setting, we need to take this pattern into account.

A similar graphical analysis for Airbnb prices and occupancy rates is not informative. The reason is that Airbnb capacity never reaches its limits. However, Figure 3 suggests that Airbnb prices do not react as strongly to changing demand as hotel prices do. Instead, Figure 1 indicates that Airbnb listings may rather react to changing demand by adjusting supply. To assert whether this relationship is true, we propose to analyze the joint movement of the deviation of Airbnb occupancy and supply from long-term trends. To calculate long-term Airbnb occupancy and supply, we calculate a 14-day moving average for both. Next, for each date, we calculate the deviation of actual occupancy and supply from these moving averages. These deviations can be interpreted as short-term fluctuations in Airbnb occupancy and supply, net of long-term trends. Figure 4b shows a scatterplot of this residual occupancy and supply with a linear fit. Each point represents one date. The positive slope of the linear fit suggests that as occupancy (i.e. demand) for Airbnb increases in the short run, the supply of Airbnb also tends to be larger. Thus, this figure is in line with Airbnb hosts flexibly reacting to changing market situations and, consequently, Airbnb supply being more flexible than hotel supply.

Figure 4a suggests that capacity constraints of hotels play an important role in the market for short-term accommodations. In the final set of descriptives, we assess the prevalence of situations in which these capacities are constrained. Figure 5 shows the city-wide hotel occupancy ratio over time in the solid black line. The horizontal dashed line marks an occupancy ratio of 90 percent which we define as a high occupancy state. There are several dates during the year when the black line is above this 90 percent cut-off. Farronato and Fradkin (2018) find that consumers profit most from the presence of Airbnb during these high occupancy states. However, in Figure 5, we further document that localized high occupancy states occur far more frequently than city-wide high occupancy states. The dotted line shows the share of districts in which hotel occupancy exceeds 90 percent at each date. For many



(a) Prices and occupancy ratios (OR) for four-star hotels in the first district

(b) Short-term supply and occupancy for Airbnb listings net of long-term trends with linear fit

Figure 4: Occupancy rate and hotel and Airbnb behavior

dates when the city-wide hotel capacity does not seem to be constrained, it actually is in individual districts of the city. This insight implies that demand variation is not distributed uniformly across the city. If, additionally, consumer preferences are correlated within districts in the city, ignoring this within-city variation in occupancy rates, prices, and demand, likely distorts the estimation. One main advantage of our data compared to prior research is that it allows us to exploit the local variation observed in the data.

The capacities of Airbnb never really reach their limits. However, specifically when hotels are capacity constrained, Airbnb listings can serve as a valuable alternative for consumers (Farronato and Fradkin, 2018). Figure 6 shows the relationship between the occupancy ratio of hotels and Airbnb listings in an exemplary district. Note that the two axes have different scales. The vertical dashed line marks a 90 percent occupancy ratio of hotels. Each point represents one date. This figure provides two main insights. First, there is a positive correlation between hotel and Airbnb occupancy. Figure 2 already suggests such a correlation for absolute demand for hotels and Airbnb listings; Figure 6 confirms this pattern for occupancy ratios as well. Second, when hotel occupancy is high, Airbnb occupancy seems to be disproportionately higher as well. The cloud of points in the top right corner of the

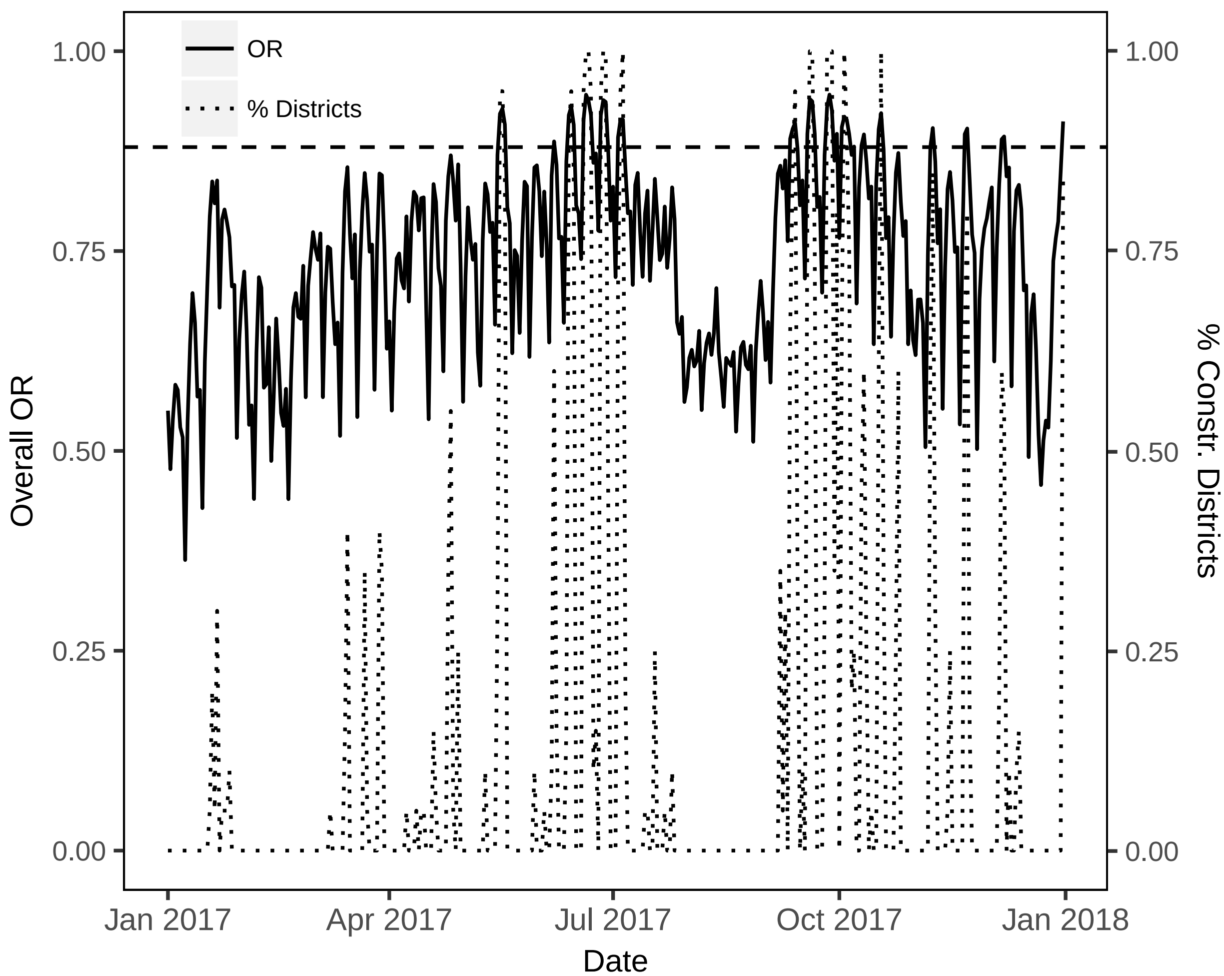


Figure 5: City-wide hotel occupancy ratio (OR) and percentage of districts in which hotel occupancy ratio is above 90 percent. The horizontal dashed line marks an occupancy ratio of 90 percent.

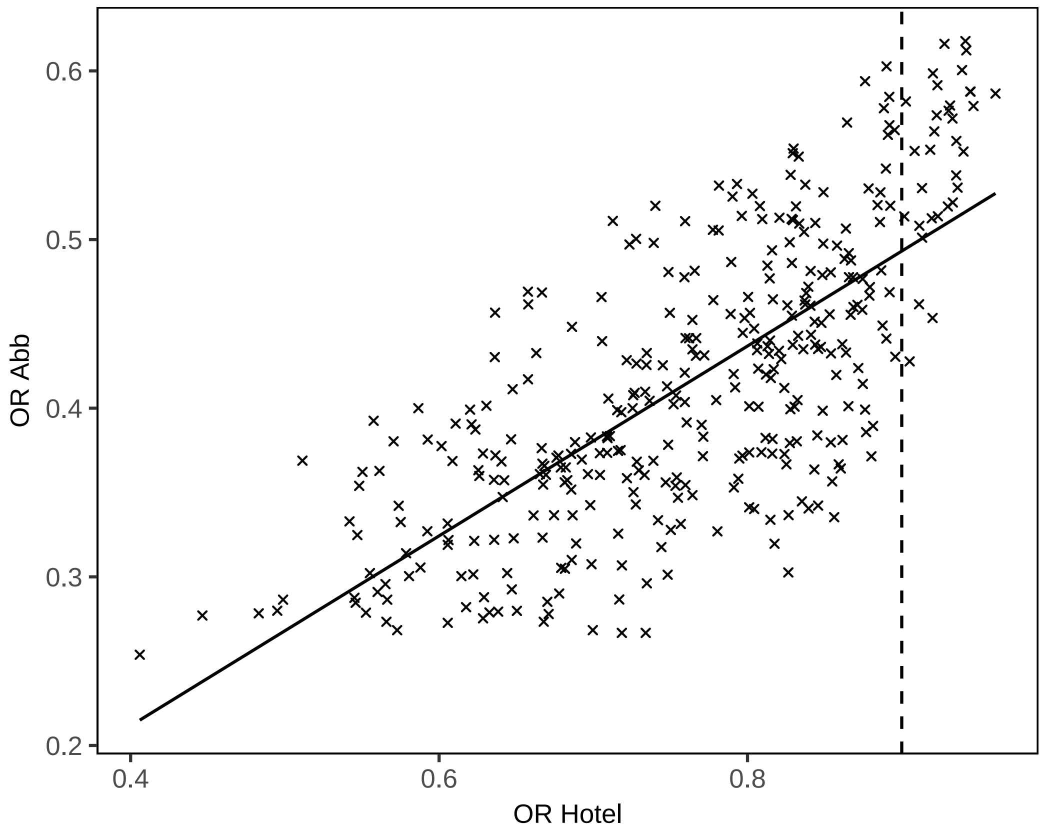


Figure 6: Occupancy ratio (OR) of hotels and Airbnb listings in first district

plot suggests that substitution from hotels to Airbnb listings might be more relevant when hotels are capacity constrained. This insight indicates that these high occupancy states are particularly important when considering how consumers profit from the presence of Airbnb.

3.2 Model

The evidence in Section 3.1 suggests that demand for short-term accommodations in Paris varies not only over time but also between different districts in the city. Furthermore, while some substitution between Airbnb and hotels takes place (compare Figure 6), demand does seem to be somewhat segregated between the two types of accommodation. Therefore, we propose to base the demand estimation on a two-level nested logit model. The first nesting level represents a choice between different districts of the city. Figure 5 suggests that much of the variation in demand over time is not distributed uniformly across the city but is localized in individual districts. Defining the choice between districts as the first nesting level allows for preferences to be correlated among alternatives within each district and, therefore, can capture the important role that within-city geography plays in the demand for accommodations. The second nesting level represents the choice between Airbnb listings and hotels within each district. While the descriptive evidence suggests that geography is a major factor in the demand for accommodations, this second nesting level allows for some additional segmentation of demand along the types of accommodation.

We next describe the model more formally. We define each date as an individual market t. Let J_t denote the set of available choices in market t. This choice set is mostly constant over time and consists of the products described in Section 2.3 (combinations of district, accommodation type, and quality category). However, in some markets, some Airbnb types are missing in certain districts. Each consumer chooses either one of the products in J_t or the outside good. Let $d = 0, ..., D$ denote the district nests, where $d = 0$ is the outside good. Further, define H_d as the subnests within each nest d.

The two-level nested logit model implies the following choice probability for product j, located in subnest h_d in nest d:

$$P_{jt} = P_{jt|h_d} \times P_{h_d|d} \times P_d \tag{1}$$

$P_{jt|h_d}$ denotes the choice probability of product j conditional on the consumer choosing subnest $h_d \in H_d$. $P_{h_d|d}$ denotes the probability of choosing subnest h_d conditional on choosing nest d. P_d denotes the probability of choosing nest d. The (conditional) probabilities in Equation (1) can be rewritten as (we follow the notation introduced by Verboven (1996)):

$$P_{jt} = \frac{e^{\delta_{jt}/(1-\sigma_1)}}{e^{I_{hdt}/(1-\sigma_1)}} \times \frac{e^{I_{hdt}/(1-\sigma_2)}}{e^{I_{dt}/(1-\sigma_2)}} \times \frac{e^{I_{dt}}}{\sum_{d=0}^{D} e^{I_{dt}}}\,, \tag{2}$$

where I_{hdt} and I_{dt} represent the expected utility from choosing an alternative in subnest h_d

and nest d, respectively. These expressions are given by:

$$\begin{aligned} I_{h_d t} &= (1-\sigma_1) ln\Big(\sum_{j \in h_d t} e^{\delta_{jt}/(1-\sigma_1)}\Big) \text{ and} \\ I_{dt} &= (1-\sigma_2) ln\Big(\sum_{h_d \in H_{dt}} e^{I_{h_d t}/(1-\sigma_2)}\Big) \,. \end{aligned} \tag{3}$$

The parameters σ_1 and σ_2 govern the correlation of preferences between alternatives in the same subnest and nest, respectively. The parameters must fulfill $0 \le \sigma_2 \le \sigma_1 \le 1$ to be consistent with random utility maximization of the consumers (McFadden et al., 1977).

A larger value of σ implies stronger correlation in the unobserved preference components. The inequalities therefore imply that the preferences across alternatives in the same subnest must be weakly more strongly correlated than preferences across alternatives in the same nest, but different subnests. Note that, in principle, the parameters σ_1 and σ_2 can vary across different nests. A fully flexible model would require estimating 60 parameters: 20 parameters for each upper level nest (one for each district) and 40 parameters for each lower level subnest (one for hotels and one for Airbnb in each district). To facilitate estimation, we restrict σ_1 and σ_2 to be identical across nests and subnests.

δ_{jt} is the mean utility of alternative j in market t, which we define as follows:

$$\delta_{jt} = x_{jt}\beta - \alpha p_{jt} + \epsilon_t + \epsilon_{jt} \,. \tag{4}$$

x_{jt} is a vector of observable product characteristics, p_{jt} denotes the price, ϵ_t is a market-specific fixed effect and ϵ_{jt} is a product-specific time varying error term that is potentially correlated with the price. We assume that vector x_{jt} is uncorrelated with ϵ_{jt}. In our analysis, x_{it} contains quality category fixed effects, a constant, and, in some specifications, time fixed effects, high-capacity districts fixed effects, and the log of exogenous Airbnb supply. All of these variables are fixed in the short term and cannot react to changing demand. Therefore, we believe that the assumption that x_{jt} is exogenous, is reasonable.

3.3 Estimation

We estimate the parameters α, β, σ_1, and σ_2 using the market share inversion technique outlined in Berry (1994) and Berry et al. (1995). Verboven (1996) derives the inverted market shares for a two-level nested logit model, which yields the following estimation equation:

$$ln(s_{jt}/s_{0t}) = x_{jt}\beta - \alpha p_{jt} + (\sigma_2 - \sigma_1)ln(s_{jt}/s_{hdt}) + \sigma_2 ln(s_{jt}/s_{dt}) + \epsilon_t + \epsilon_{jt} \tag{5}$$

s_{jt} is the share of product j in market t. s_{0t} denotes the share of the outside good. s_{hdt} is the cumulative share of products in subnest h_d and s_{dt} is the cumulative share of products in nest d.

To determine the market shares, we require an estimate of the market size, i.e. a measure of the number of people looking for short-term accommodation in Paris. We compute the market size using worldwide Google trends data for the keyword "hotels Paris" and "Airbnb Paris." The data span the last four weeks of 2016 and all weeks in 2017. We add up both trends to obtain a measure for the total search interest for both accommodation types. We normalize the aggregate trend by setting its average value equal to the average combined number of hotel and Airbnb rooms in 2017. As a result, one unit of the combined Google trends data corresponds to 1100 searches.

The dashed line in Figure 7 shows our measure for the search activity. Our measure suggests that, on average, there are 134,000 daily searches globally for either Airbnb listings or hotels in Paris in 2017. The measure ranges from 107,000 to 165,000. Based on these search numbers, we calculate our main market size measure for each date as the average of the searches in the four weeks prior to that day. The solid line in Figure 7 presents this measure. This moving average captures the idea that travelers plan their trip in advance. However, our results are robust to using the number of searches at each date as the market size (i.e. the dashed line in Figure 7).[21]

[21] Table 8 in Appendix A.4 reports the results of this robustness check.

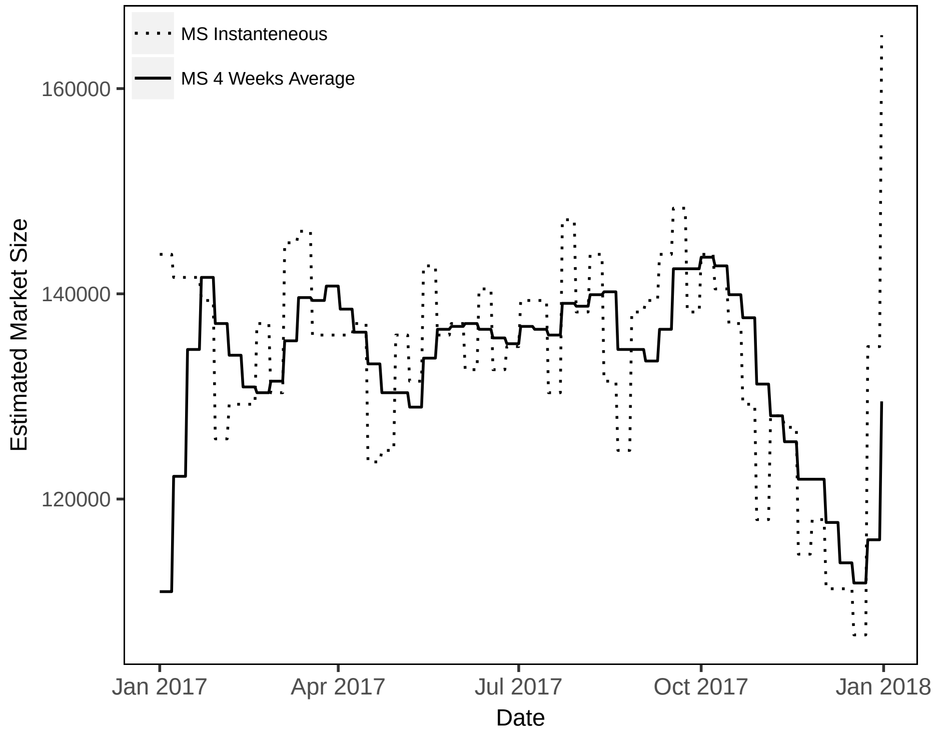


Figure 7: Market size (MS) measures

3.4 Identification

There are two main concerns for the identification of the model parameters. The first is the standard concern in demand estimation: There are three endogenous variables on the right hand side of Equation (5): the price and the two share terms. Therefore, we need at least three instrumental variables to consistently estimate Equation (5). The second concern is that our model does not capture differences in the capacities of the products. In its basic form, Equation (5) does not account for the number of individual Airbnb listings or hotels that are underlying each of the aggregated products. Ignoring changes in these numbers might bias our results.

Instrumental variables One main concern regarding the endogeneity of prices and market shares are time-varying localized demand shocks. As discussed in Section 3.1, the different districts are subject to short-term demand variation that is heterogeneous in location and intensity over time.

Demand shocks enter ϵ_{jt} and are positively correlated with prices. Without appropriate instrumental variables, this will result in an underestimation of the magnitude of the true price parameter. Note that the localized demand shocks will also lead to biased parameter estimates of σ_1 and σ_2 since any unobserved demand shock that affects s_{jt} will necessarily also affect the share terms on the right hand side of Equation (5). Therefore, we require instrumental variables that are independent from local, short-term demand variation and correlated with prices and market shares.

Our instrumental variable strategy relies on the insight that the price setting behavior of a given accommodation is dependent on the room capacity of its nearby rivals. *Ceteris paribus*, higher room capacity of nearby competitors should reduce the ability to charge higher prices.

Therefore, we propose to construct instruments based on the room capacity of similar products (i.e. the same accommodation type and quality category) in other districts. In this measure, further away competitors should be weighted less. To formalize this idea, consider a product j, in market t. Let A_{jt} be the set of products of the same type and quality in market t. With 20 districts, this set can contain a maximum of 20 elements. We construct the instrument for product j as:

$$z_{jt} = \sum_{\substack{l \in A_{jt} \\ l \neq j}} \frac{c_{lt}}{e_{lj}^2}\,, \tag{6}$$

where c_{lt} is the capacity of product l and e_{lj} is the Euclidean distance in kilometers between products j and l.[22] Thus, the instrument is the weighted sum of the rival capacity, with weights that are inversely related to the distance to the rival.

For z_{jt} to be a valid instrument, it must be uncorrelated with short-term variation in demand. Figure 1 suggests that this is likely the case when calculating the measure for hotels, as hotel demand is almost constant over time. However, for Airbnb supply, the descriptives presented in Section 3.1 suggest that the capacity does respond to demand shocks. Therefore,

[22] We use the centroids of each district to calculate the distances.

we need a measure of the *exogenous* supply of Airbnb listings to use in the calculation of Equation (6) for Airbnb listings.

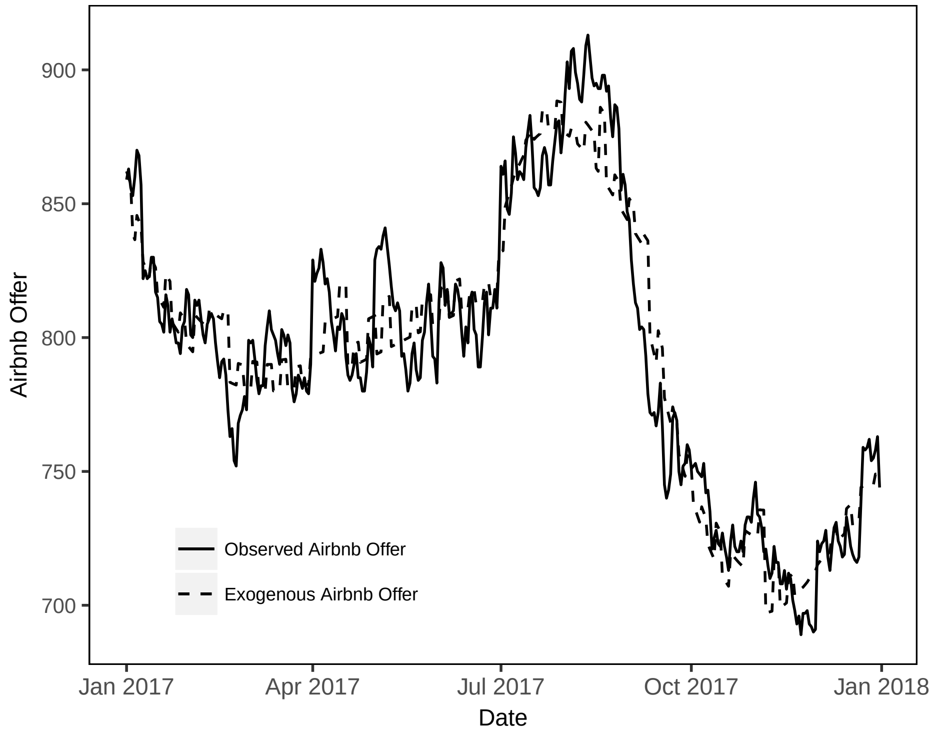


Figure 8: Observed and exogenous Airbnb offer for selected Airbnb product

For this purpose, we propose to predict the Airbnb supply for each product using explanatory variables that are plausibly exogenous with regard to short-term variation in demand. More specifically, we predict Airbnb supply based on a quartic time trend and a measure of leisure-related outgoing car traffic from Paris. The quartic time trend flexibly captures changes in Airbnb supply in the city due to long-term trends that should be independent from short-term and localized variation in demand. To obtain a measure of leisure-related outgoing car traffic, we use data that measure outgoing car traffic at main highway exits around Paris.[23] To distinguish between business traffic and leisure-induced traffic, we construct a measure of weekend and holiday traffic. On the day before the weekend or the holiday (i.e. on Fridays for each weekend), we take the amount of outgoing traffic and subtract the av-

[23]The data are available at https://opendata.paris.fr/explore/dataset/comptages-routiers-permanents/information/?disjunctive.libelle&disjunctive.etat_trafic&disjunctive.libelle_nd_amont&disjunctive.libelle_nd_aval (accessed: July 20, 2020).

erage amount of outgoing traffic during the weekdays of the preceding week.[24] We interpret this excess traffic as travelers leaving the city for the weekend or for the holidays. For the estimation, we assign this excess value to the entire respective weekend or holiday. We let the measure be zero for weekdays. Therefore, weekday variation in Airbnb supply is only explained by long-term trends. The procedure is applied for each Airbnb product separately to allow for product-specific variation of the exogeneous offer. Figure 8 shows the observed and predicted Airbnb offer for a selected Airbnb product. The deviations from the long-term trend are caused by the variation in the traffic data. Having predicted Airbnb supply using the quartic time trend and the leisure-induced outgoing car traffic, we use these predicted Airbnb numbers to calculate Equation (6) for all Airbnb products.

With this procedure, we obtain measures of Equation (6) that only capture same-type capacities in the city. However, the capacities of products of the other type with similar quality are likely important for price setting and market shares as well. To capture this, we assign the values calculated in Equation (6) to products of similar quality but different type in the same district. For example, say we calculate z_{jt} for the category-four Airbnb product in district one. We would then assign this value also to four- and five star hotels in district one.[25] As a result, each observation has two instruments: The capacity of same type and quality products in other districts as well as the capacity of different type and similar quality products in other districts. Finally, we interact each of these two instruments with a fixed effect for hotels. Through this interaction, the instruments can affect hotels differently than Airbnb listings. This choice is motivated by the analysis in Section 3.1, which reveals differences in the response of Airbnb and hotels to short-term variation in demand.

We generate an additional instrument using the capacity of accommodations of the other type in the same district. For example, for Airbnb listings in a given district, we use the

[24]The summer holiday peak in Airbnb offer is an exception. We do not use the outgoing traffic to capture this peak. Instead, we interact the quartic time trend with a dummy variable that captures the period of the summer holidays.

[25]Because we have four Airbnb and five hotel quality categories, we assign the values of the highest Airbnb category to both the four as well as the five star hotels.

total hotel capacity in that district. For hotels, we use the total Airbnb capacity. Again, we interact this instrumental variable with a fixed effect for hotels.

As noted above, the capacity of hotels is nearly constant over time. To generate temporal variation in the instruments based on hotel capacity, we interact all the above instruments with the market size. The identifying assumption is that the market size is independent of the exact location of demand shocks. The intuition behind this assumption is that the total market size is an aggregate of the location-specific demand. Therefore, variation in demand for specific locations in the city does not affect total demand to travel to Paris much and is offset on average. The descriptive analysis in Figure 5 shows that periods of district-level hotel capacity constraints becoming binding are much more frequent than periods of city-wide hotel capacities being at their limits. This insight is in line with our assumption.

Another issue regarding the validity of the market size as an instrument is its relationship with the overall price level in the city. The market size reflects the number of individuals who are considering travel to the city at a specific point in time. Long-term variation in average accommodation prices across different tourist destinations most likely affect the market size for each of these destinations. However, for us to use the market size as an instrument, we only require that it is uncorrelated with short-term variation in prices. If most individuals traveling to Paris are somewhat inflexible in the dates of travel (this would be the case e.g. for business travelers or tourists planning their trips in advance), then short-term price variation should not affect the total market size.

The reverse mechanism that hotels anticipate periods of high demand and set higher prices accordingly is more problematic. Consumers' responses to relative price differences within the city should remain unaffected by the overall shift in prices. This suggest that the variation in relative shares within the city as a response to variation in price differentials within the city should allow us to consistently estimate the price parameter.[26]

[26]We perform a simulation study where we assess the validity of the instrument. In our simulation, four products (which can be understood as hotels) adjust their prices in response to overall market size and product-specific shocks. Consumers' choice probabilities follow a simple logit model. Other than prices, products are differentiated by one exogenous observed characteristic. As an instrument we use the average

Accounting for capacity Our model does not capture changes in the number of hotels or Airbnb listings underlying each of the aggregated products. Accounting for differences in these numbers might be relevant for our results. However, the descriptives presented in Section 3.1 suggest that the Airbnb and hotel supplies behave fundamentally differently. Consequently, accounting for differences in the actual number of rooms underlying the products requires different approaches for each accommodation type. For hotels, the main concern is that differences in hotel capacities might be driving part of the differences in market shares. Hotels with low capacity tend to have lower market shares. Therefore, market shares might reflect capacity constraints rather than consumer preferences. Because hotel capacities do not change much in our sample, we propose to capture these differences by including fixed effects for hotels in districts with a high capacity of four- and five-star hotels.[27] We define a district as a high-capacity district if it contains more than 2,000 four- and five-star hotel rooms.[28] This criterion applies to six districts. We include a separate fixed effect for each of these districts. We let the fixed effects be zero for all Airbnb listings.[29]

The issue is different for Airbnb capacities. Because the number of Airbnb listings changes over time, a fixed effect would not be sufficient to capture the effect of these changes. Instead, we follow Ackerberg and Rysman (2005) and include the log of the Airbnb capacity (we add one to allow for capacities of zero) in the estimation. Section 3.1 shows that Airbnb capacity likely reacts to short-term changes in demand. Therefore, we only include the log of the

value of the exogenous product characteristic of rivals (classical BLP instrument) and interact it with the market size. The resulting 2SLS estimator produces a consistent estimator. Details of the simulation study are presented in Appendix A.2, where we also compare the performance of the proposed instrumental variable with the classical BLP instrument without interaction with the market size.

[27]We only focus on four- and five-star hotels because these make up a majority of overall hotel demand. High-quality hotels with large capacities are more likely to host events such as conferences. Furthermore, four- and five-star hotels also reflect differences in hotel capacity across districts well. In districts in which there is a large capacity of four- and five-star hotels, overall room capacity tends to be higher. As a potential alternative to our approach, de Palma et al. (2007) propose a method to account for capacity constraints in demand estimation. However, their method is only developed for a simple multinomial logit. Extending this method to more flexible demand models is an interesting avenue for future research.

[28]Effectively, a district is defined as high capacity if it has a four- and five-star hotel capacity of 2,629 or more hotels which is the smallest of the district capacities above 2,000. A capacity of 2,629 corresponds approximately to the 75 percentile of the distribution of district-wide four- and five-star hotel capacities.

[29]In Figure 14 of Appendix A.1, we show the distribution of hotel and Airbnb capacity over districts.

exogenous Airbnb capacity that we also use to calculate Equation (6) for Airbnb listings. For hotels, we let this variable be zero. These additional variables allow us to account for changes in the market share that might be driven by unobserved changes in the underlying number of rooms and listings. A caveat of these additional variables is that they lack a clear structural interpretation. Therefore, we also report results without these additional variables.[30]

4 Results

4.1 Parameter Estimates

Table 1 shows the result of estimating different specifications of Equation (5). Estimation is performed using two-stage least squares, the reported standard errors are heteroskedasticity robust.[31]

Column (1) reports the results for a simple logit estimation without instruments. The positive price coefficient suggests that it suffers from an upward bias due to unobserved factors increasing both demand as well as prices. The estimated quality category fixed effects are also counter-intuitive: For instance, five-star hotels are valued lower than two-star hotels.

Columns (2) and (3) show the results for a simple logit specification and the two-level nested logit model of Equation (5) when using the instruments presented in Section 3.4. Column (4) includes market-specific fixed effects. Column (5) further includes the natural logarithm of the exogenous Airbnb offer and district-specific fixed effects for all hotels located in a district with a high capacity of four- and five-star hotels. In all of these instrumented specifications, the price coefficient is statistically significantly negative and appears robust

[30] An alternative way to assess the sensitivity of our results to the varying Airbnb offer is to restrict the estimation sample to only contain observations between February 1st and July 1st. As Figure 1 reveals, the Airbnb offer varies less during this period. We present the results of the corresponding analysis in Table 9 of Appendix A.4. Our main results are not affected when restricting the sample period.

[31] We report the first-stage F-statistics in Table 7 of Appendix A.3. We use the same set of instruments for each endogenous variable. The F-statistics suggest that the instruments are not weak.

Table 1: Parameter estimates

	(1) Logit	(2) Logit (IV)	(3) NL	(4) NL	(5) NL
Price	0.002***	−0.008***	−0.006***	−0.006***	−0.005***
	(0.0001)	(0.0004)	(0.0003)	(0.0003)	(0.0003)
σ_1			0.758***	0.657***	0.717***
			(0.017)	(0.014)	(0.018)
σ_2			0.256***	0.228***	0.207***
			(0.010)	(0.009)	(0.006)
Airbnb category 2	0.290***	0.501***	0.068***	0.132***	−0.019*
	(0.041)	(0.044)	(0.021)	(0.023)	(0.011)
Airbnb category 3	0.421***	0.924***	0.280***	0.379***	0.078***
	(0.040)	(0.047)	(0.027)	(0.027)	(0.018)
Airbnb category 4	0.261***	1.586***	0.774***	0.920***	0.456***
	(0.042)	(0.067)	(0.044)	(0.044)	(0.038)
1-star hotel	−0.539***	−0.148***	0.698***	0.619***	2.553***
	(0.048)	(0.052)	(0.034)	(0.030)	(0.066)
2-star hotel	0.740***	1.319***	0.968***	1.032***	2.712***
	(0.041)	(0.048)	(0.025)	(0.025)	(0.093)
3-star hotel	2.067***	2.922***	1.463***	1.674***	3.232***
	(0.041)	(0.055)	(0.041)	(0.040)	(0.121)
4-star hotel	1.644***	3.182***	1.822***	2.045***	3.505***
	(0.043)	(0.075)	(0.050)	(0.050)	(0.126)
5-star hotel	−0.210***	4.326***	3.206***	3.482***	4.468***
	(0.066)	(0.176)	(0.114)	(0.111)	(0.165)
Log Airbnb offer					0.351***
					(0.016)
Constant	−6.470***	−6.031***	−4.157***	−4.307***	−6.319***
	(0.039)	(0.044)	(0.045)	(0.070)	(0.128)
Market FEs	*N*	*N*	*N*	*Y*	*Y*
N	33,504	33,504	33,504	33,504	33,504
Adjusted R^2	0.377	0.133	0.652	0.700	0.862

Notes: The lowest Airbnb quality category serves as the base category. *, **, *** indicate statistical significance at the ten, five, and one percent level, respectively. F-statistics are shown in Appendix A.3.

across specifications.

Generally, consumers seem to value hotels more highly than Airbnb listings. Our results suggest that Airbnb listings of the highest quality are valued similarly to two-star hotels. Note that in column (5), the Airbnb and hotel fixed effects are not readily comparable because we include high-capacity district fixed effects that apply to hotels only while the log Airbnb offer applies only to Airbnb listings. The coefficient for the logarithm of the exogenous Airbnb offer is statistically significantly different from zero. Ackerberg and Rysman (2005) propose a structural interpretation for this coefficient: The result implies that consumers value higher availability of Airbnb listings because, as a result, accommodations are more likely to be in line with their geographic preferences.[32]

The estimate for σ_1 suggests strong taste correlation between products in the same subnest. Recall that each subnest is an accommodation type (hotel or Airbnb listing) in a given district. Furthermore, the estimate for σ_2 indicates taste correlation between products in the same district. However, the correlation between products in the same nest (i.e. in the same district) is smaller than for products in the same subnest (e.g. Airbnb listings in the same district). These results are in line with localized competition between Airbnb and hotels. However, localized competition within hotels and Airbnb listings in the same district is stronger than the competition between hotels and Airbnb listings in the same district.

4.2 Elasticities

Table 2 reports the average estimated own- and cross-price elasticities for different Airbnb and hotel quality categories based on the estimates of column (5) in Table 1.[33] Demand seems to be elastic for all quality categories except for the two lowest Airbnb quality categories.

[32]For better exposition of the main results, we do not report the estimated coefficients for the high-hotel-capacity district fixed effects described in Section 3.4. All of these coefficients are statistically significantly positive, as expected. The high-hotel-capacity fixed effects lack the same structural interpretation of the logarithm of the Airbnb offer because high capacity does not necessarily imply large variety in the location of hotels within a district: In the extreme, a single hotel could make up the majority of the hotel capacity in a district.

[33]The formulas for the own- and cross-price elasticities are presented in Appendix A.5.

The own price elasticity tends to increase with the quality. Generally, the demand for hotels seems to be more elastic than the demand for Airbnb listings. However, the two highest Airbnb quality categories have comparable own-price elasticities to those of three- and four-star hotels, respectively.[34]

Table 2: Average estimated demand elasticities

			Cross-price elasticities		
Type	Category	Own-price elasticity	Same district & Type	Same district	Other district
Airbnb	1	-0.5095	0.2548	0.0091	0.0007
	2	-0.9583	0.1744	0.0057	0.0004
	3	-1.3400	0.2387	0.0077	0.0006
	4	-2.4661	0.6033	0.0178	0.0014
Hotels	1	-1.3769	0.0669	0.0053	0.0005
	2	-1.6417	0.1310	0.0109	0.0011
	3	-1.6144	0.6378	0.0538	0.0054
	4	-2.6882	0.7415	0.0629	0.0060
	5	-7.4321	1.2201	0.1111	0.0120

Given the estimates of σ_1 and σ_2, cross-price elasticities are highest for products in the same subnest. Cross-price elasticities have intermediate values for products that only share the same nests but are in different subnests. Finally, cross-price elasticities are lowest for products in different nests.

For example, Table 2 reveals that a one percent price increase for five-star hotels leads to an average 1.2 percent increase in demand for hotels in the same district. The same price increase results in a demand increase of only 0.1 percent for Airbnb listings in the same district. The demand for accommodations in other districts only increases by 0.01 percent following the price increase.

[34] We report the semi-elasticities in Appendix A.6. As opposed to the elasticities, the own-price semi-elasticities tend to decrease with the quality category, although the pattern is not as clear cut as for the own-price elasticities.

5 Airbnb, Consumer Welfare, and Hotel Performance

In this section, we assess the impact of Airbnb on hotel revenues and consumer welfare. There are two main channels through which Airbnb benefits consumers. First, Airbnb increases the choice variety. Second, Airbnb's presence in the market is likely to reduce hotel prices. To quantify these effects of Airbnb, we analyze counterfactual scenarios in which Airbnb does not exist and compare them to a scenario with Airbnb. In order to determine hotel prices in these counterfactual scenarios without Airbnb, we need to address the supply side of the market more closely. In the remainder of this section, we first present our supply-side model. Based on this model, we can simulate the price response of hotels to Airbnb leaving the market. This allows us to compare the loss in consumer surplus when Airbnb leaves the market *and* hotels adjust prices to a scenario in which hotels do not adjust prices when Airbnb leaves the market. Additionally, modelling the price response of hotels in the counterfactual scenario where Airbnb is not present in the market allows us to assess the impact of Airbnb on hotel revenues.

The standard approach to quantifying the impact of Airbnb would be to simulate counterfactual scenarios without Airbnb and compare the simulated outcomes to the observed outcomes in the data. However, note that the observed data are characterized by location-specific demand shocks that our model does not account for.[35] Hence, our simulation analysis is not able to incorporate these demand shocks. Therefore, we do not compare simulated outcomes (which do not incorporate unobserved short-term localized demand variation) with observed outcomes (which do include unobserved short-term localized demand variation). Instead, we simulate market outcomes not only for the scenarios without Airbnb but also for a benchmark scenario including Airbnb. Subsequently, we compare hotel performance and

[35] In fact, our identification strategy builds on finding instruments that are orthogonal to location-specific demand shocks. Another approach would be to estimate market-district fixed effects directly. These fixed effects would capture short-term district-specific variation in demand. Subsequently, we could use the estimates in the counterfactual simulations. Unfortunately, this approach appears to be too demanding of the data. Particularly, note that with market-district fixed effects, the number of parameters grows with the sample size, which gives rise to dimensionality issues.

consumer welfare in the scenarios without Airbnb to this benchmark scenario. Comparing simulation outcomes can be seen as studying the impact of Airbnb while abstracting from location-specific demand shocks.

To simulate scenarios without Airbnb, we only require a model to simulate hotel behavior absent Airbnb. However, to simulate the benchmark scenario, we also need to capture Airbnb supply. Therefore, we first discuss how we incorporate Airbnb supply in our simulations. Then, we describe how we model hotel supply. Subsequently, we present how well our simulation of the benchmark scenario fits the observed data. Finally, we report the results of our simulations.

5.1 Airbnb Supply

To include Airbnb supply in our simulation, we propose to use the exogenous measure of Airbnb supply discussed in Section 3.4. This exogenous Airbnb offer allows us to include Airbnb supply in a direct and simple way and is consistent with abstracting from local, short-term variation in demand in our simulations.[36] For Airbnb prices in our simulation of the benchmark scenario, we use the observed Airbnb prices. The descriptive analyses in Section 3.1 suggest that Airbnb supply reacts to short-term changes in demand by adjusting offered quantities rather than prices. Therefore, short-term variation in demand should not affect these observed Airbnb prices much.

5.2 Hotel Supply-Side Model

In this section, we describe our approach to model hotels' equilibrium prices. Conventionally, discrete choice models are extended to include the supply side by assuming a mode of competition between firms and deriving firms' profit-maximizing first order condition. This

[36] Farronato and Fradkin (2018) model Airbnb hosts as price-takers who increase supply in response to higher prices. Since our model does not incorporate localized demand shocks, Airbnb supply predicted by an entry model would not adapt to demand shocks and therefore remain largely constant within a location across time. Therefore, the insight from modeling Airbnb entry in our context would be limited.

first order condition can then be used to derive marginal costs based on estimated demand elasticity and observed prices (Berry, 1994; Verboven, 1996). Together with the parameter estimates, these estimated marginal costs then allow for simulating prices in counterfactual scenarios.

In our application, this standard approach is problematic for two main reasons. First, as Figure 5 suggests, hotel capacity constraints play an important role in this market. Discrete choice models, however, typically assume unconstrained capacities on the supply side. When estimating the demand model, these constraints are somewhat alleviated because the demand parameters fit the theoretical market shares to the observed market shares, which, by definition, adhere to the capacity constraints. However, when simulating counterfactual situations, nothing restricts the resulting market shares from exceeding the actual capacities.

Second, optimal pricing of hotels likely requires internalization not just of the own capacities but also those of the competitors. This cross-dependency between competitors makes finding an analytical solution for the market equilibrium complex.

Therefore, we propose an approach to simulating hotel prices and quantities in counterfactual equilibria using a somewhat more heuristic, iterative approach. There are several components to our approach. The first component we need are hotels' optimal prices as a function of their occupancy rates. While in reality, this function might also rely on competitors' occupancy rates, we propose to use the observed, empirical relationship between hotel prices and own occupancy as a reduced form approximation. The underlying assumption is that prices in the data are the result of optimal pricing by hotels. Assuming that the relationship between occupancy rates and optimal prices remains fixed without Airbnb, we can use the observed empirical relationship between prices and occupancy rates to simulate optimal hotel prices.

To estimate the function of optimal hotel prices dependent on occupancy ratios, we regress the observed prices on the corresponding occupancy ratios. Motivated by the insights from Figure 4a, we assume a kinked linear relationship between occupancy ratio and prices, with

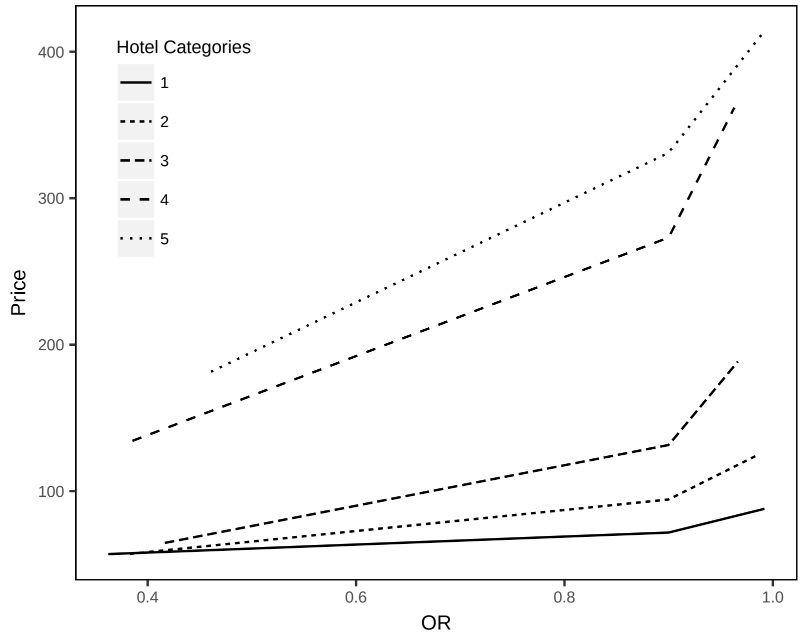


Figure 9: Estimated pricing function for different hotel types in an exemplary district

a kink at an occupancy ratio of 90 percent. We estimate a separate pricing function for each hotel product. Figure 9 shows the resulting estimated optimal prices as a function of the occupancy ratio for different hotel types in an exemplary district. The results suggest that higher quality hotels are more expensive and price more steeply (in absolute terms) when approaching their capacity constraints. Denote this estimated optimal pricing function as $\hat{p}^*(q)$.

The second component we need in any counterfactual simulation is equilibrium demand. In principle, we can calculate market shares given parameter estimates and some price vector $\boldsymbol{p}$ using Equation (2). However, these counterfactual market shares do not necessarily adhere to hotel capacity constraints. We next propose an iterative method to allocate predicted demand to hotels while taking into account their capacity constraints.

For now, let the price vector $\boldsymbol{p}$, the other product characteristics, as well as demand parameters be fixed. Using Equation (2), we can then calculate predicted occupied rooms for each product. Denote the vector of predicted occupancy as $\boldsymbol{q_\tau}(\boldsymbol{J_\tau}, \mathbb{d}_\tau)$, where $\boldsymbol{J_\tau}$ denotes the set of products used as choice set in Equation (2) in iteration τ and $\mathbb{d}_\tau$ denotes the number of travelers. We make the dependence of $\boldsymbol{q_\tau}$ on $\boldsymbol{J_\tau}$ and $\mathbb{d}_\tau$ explicit because these are the only components of demand that change in each iteration. Note that the dimension of vector $\boldsymbol{q_\tau}$ equals the number of products in $\boldsymbol{J_\tau}$ and can change with each iteration. For the first iteration $\tau = 1$, let $\boldsymbol{J}_1$ be the set of all products and $\mathbb{d}_1$ be the full market size. In each iteration τ, we proceed as follows:

1. Calculate $\boldsymbol{q_\tau(J_\tau, \mathbb{d}_\tau)}$ using Equation (2).

2. Let q_τ^j denote the jth element of $\boldsymbol{q_\tau(J_\tau, \mathbb{d}_\tau)}$. Calculate $\tilde{q}_\tau^j = \tilde{q}_{\tau-1}^j + q_\tau^j$ for all $j \in \boldsymbol{J_\tau}$. For the first iteration, use $\tilde{q}_0^j = 0$.

3. Let k^j denote the room capacity of product j. Denote as $\boldsymbol{J_\tau^c}$ the set of products for which $\tilde{q}_\tau^j > k^j$. Denote the complementary set as $\boldsymbol{J_\tau^{\not c}} = \boldsymbol{J_\tau} \setminus \boldsymbol{J_\tau^c}$. Calculate aggregate excess demand as $\hat{Q}_\tau = \sum\limits_{j \in \boldsymbol{J_\tau^c}} (\tilde{q}_\tau^j - k^j)$.

4. Set final demand $q^j = k^j$ for all $j \in \boldsymbol{J_\tau^c}$. If $\hat{Q}_\tau = 0$ stop, else set $\boldsymbol{J_{\tau+1}} = \boldsymbol{J_\tau^{\not c}}$ and $\mathbb{d}_\tau = \hat{Q}_\tau$ and continue with step 1 of iteration $\tau + 1$.

The basic idea of the iterative process is that in each iteration, predicted hotel demand is allocated to the products up to their capacity constraints. In the next iteration, demand is predicted again but only using the excess demand from the previous iteration and those hotels whose capacities are not yet at their limits. These iterations continue until no excess demand remains.

This algorithm returns a vector of final demand $\boldsymbol{q}$ that contains the demand for each product. Note, however, that so far, the price vector $\boldsymbol{p}$ was kept fixed at some arbitrary values. Given some demand vector $\boldsymbol{q}$, we can now calculate the corresponding optimal hotel prices using $\hat{p}^*(q)$. This yields a vector of optimal prices, say $\boldsymbol{\hat{p}^*(q)}$. The quantities $\boldsymbol{q}$ were, however, obtained based on some other price vector $\boldsymbol{p}$ and are therefore not necessarily consistent with $\boldsymbol{\hat{p}^*(q)}$. Therefore, the iterative process to obtain quantities $\boldsymbol{q}$ described above needs to be nested inside an outer iteration process that determines a price vector that is consistent with the quantities. For each iteration κ of this outer loop, we then:

1. Calculate $\boldsymbol{q_\kappa}$ using price vector $\boldsymbol{p_\kappa}$ and the algorithm described above.

2. Calculate $\hat{\boldsymbol{p}}^*(\boldsymbol{q_\kappa})$.

3. If $\max|\hat{\boldsymbol{p}}^*(\boldsymbol{q_\kappa}) - \boldsymbol{p_\kappa}| \leq \epsilon$ stop, else choose $\boldsymbol{p_{\kappa+1}}$ such that $|\boldsymbol{p_{\kappa+1}} - \hat{\boldsymbol{p}}^*(\boldsymbol{q_\kappa})| \leq |\boldsymbol{p_\kappa} - \hat{\boldsymbol{p}}^*(\boldsymbol{q_\kappa})| - \delta$ and continue with step 1 of iteration $\kappa + 1$.

In step 3 of each iteration of the outer loop, we update the price vector incrementally toward $\hat{\boldsymbol{p}}^*(\boldsymbol{q})$ by an amount δ. The algorithm stops when each element in $\boldsymbol{p_\kappa}$ lies less than a threshold ϵ away from the optimal price vector $\hat{\boldsymbol{p}}^*(\boldsymbol{q_\kappa})$. In our simulation, we set $\delta = \epsilon = 10^{-5}$.[37] This double iteration procedure allows us to simulate market equilibria while taking into account the simultaneity of prices and quantities for hotels. In the end, we obtain a price vector $\boldsymbol{p}^*$ and a quantity vector $\boldsymbol{q}$ that are consistent with one another.

Our supply-side model provides an ad-hoc solution to adjust simulated market shares to comply with hotel capacity constraints. Since we do not estimate hotel costs, we cannot infer anything about hotel profits. Under the assumption of constant markups, changes in profit would be proportional to changes in revenues.

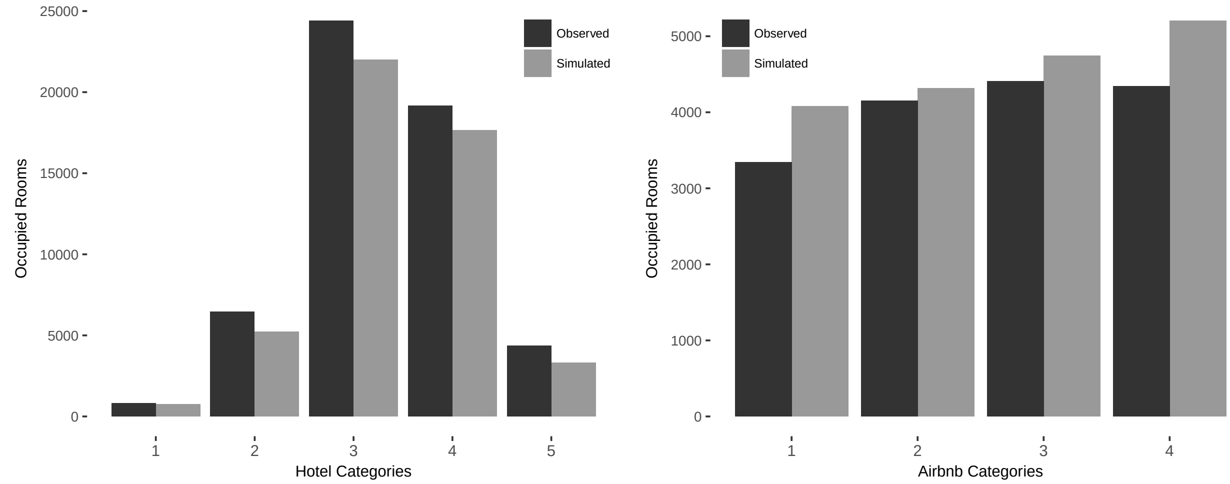
5.3 Model Fit

Before analyzing counterfactual scenarios without Airbnb, we assess how well our model matches the observed market outcomes in the benchmark scenario when Airbnb is present in the market. For the simulations, we assume that consumer utility follows the estimates reported in column (5) of Table 1 because this specification corrects for changes in Airbnb supply as well as high-hotel-capacity district fixed effects.

Figures 10a and 10b show the average daily number of occupied rooms by hotel and Airbnb quality categories, respectively. The gray bars show the quantities predicted by our

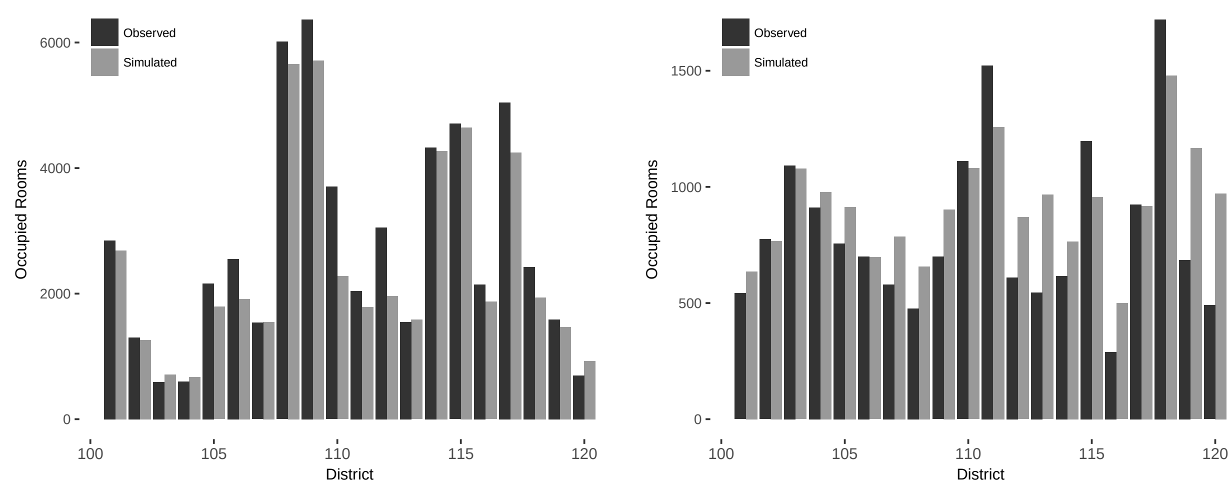
[37] In each iteration step, we could also set $\boldsymbol{p_{\kappa+1}} = \hat{\boldsymbol{p}}^*(\boldsymbol{q_\kappa})$. However, we find that the algorithm does not converge when proceeding this way. Approaching the optimal price vector incrementally leads to convergence. Varying δ and ϵ progressively leads to convergence the fastest: We start by setting $\delta = \epsilon = 1$. After convergence is reached for these initial values of δ and ϵ, we proceed with the iterative procedure setting $\delta = \epsilon = 0.1$. We continue in this fashion, reducing δ and ϵ by a factor of 10^{-1}, until convergence is reached for $\delta = \epsilon = 10^{-5}$.

model, the black bars show the quantities that we observe in the data. In general, we tend to overestimate Airbnb demand and to underestimate hotel demand. All in all, our model seems to capture the broader demand patterns reasonably well.



(a) Hotels (b) Airbnb

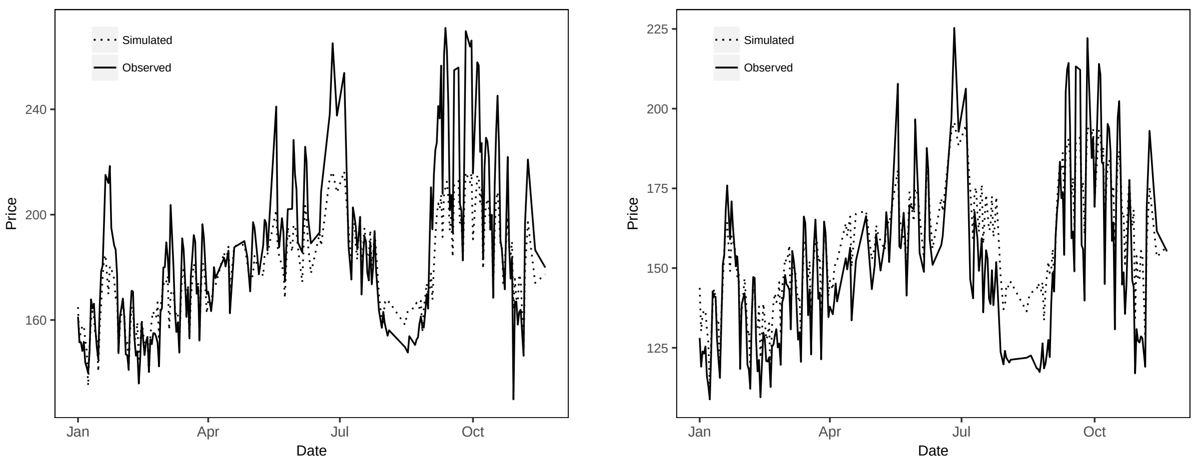
Figure 10: Simulated vs. observed demand by quality category



(a) Hotels (b) Airbnb

Figure 11: Simulated vs. observed demand by districts

Figures 11a and 11b show the average daily number of occupied rooms by district for hotels and Airbnb, respectively. Generally, we seem to simulate hotel demand well. For Airbnb listings, the simulation error appears to be more pronounced. Overall, our model



(a) Including five-star hotels

(b) Excluding five-star hotels

Figure 12: Simulated vs. observed hotel prices

captures the geographic distribution of demand well albeit with lower variance.

Turning to simulated hotel prices, our simulation manages to track the sales-weighted average daily prices, as shown in Figure 12a. However, in periods of exceptional price peaks, simulated prices tend to be lower than observed prices. This effect appears to be entirely driven by five-star hotels. If we compare the predicted sales-weighted average daily prices with the observed prices for all other hotel categories, our model matches the price data well (see Figure 12b).[38]

5.4 Simulation Results for Counterfactual Scenarios

Based on our hotel supply-side model, we can compare simulated prices and quantities for the benchmark scenario in which Airbnb and hotels are both in the market with prices and quantities when Airbnb is absent. We study two counterfactual scenarios: in the first, hotels are allowed to readjust their prices when Airbnb leaves the market. In the second scenario, hotels are not allowed to adjust their prices. Analyzing both scenarios allows us to disentangle the impact of increased product variety from the impact of hotel price

[38] Few large five-star hotels appear to attract large demand despite charging high prices (even compared to other five-star hotels). Our model does not match the observed shares for these "outlier" hotels well. Although the share of these hotels in the overall demand for accommodation is small, this mismatch heavily affects sales-weighted average prices because of the very high prices these hotels charge.

adjustments on consumer surplus. These two scenarios correspond to the two counterfactual scenarios presented in Farronato and Fradkin (2018).

For the benchmark scenario, we use the demand parameters reported in column (5) of Table 1. For the counterfactual scenarios, we assume that demand follows a one-stage nested logit model: consumers first choose the district and subsequently choose the option within the district that maximizes utility. The taste correlation between hotels of the same district is taken from our estimate for the taste correlation between accommodations in the same subnest (σ_1) in the two-level nested logit model.

Table 3 shows the simulated price changes for the different hotel types when Airbnb leaves the market. As expected, absent Airbnb, hotels increase their prices. Table 3 also displays the simulated changes in equilibrium quantities for both counterfactual scenarios: when hotels are allowed to change prices ("pr adj") and when hotels leave prices unchanged ("no pr adj"). In both counterfactual scenarios, demand is larger for all hotel categories than in the benchmark case. For most hotel categories, this quantity increase is larger if no price adjustment is allowed. Surprisingly, the reverse is true for the lowest-category hotels. This result is likely due to the different pricing patterns observed for different hotel types. Since one-star hotels increase prices less sharply in response to higher occupancy ratios, more demand is diverted to them when rivals adjust prices in response to Airbnb leaving the market. Combining these hotel price and quantity changes, we find that the average daily revenue loss for hotels due to Airbnb amounts to 1.8 million euro.

To study the consumer welfare implications of Airbnb, we compute the consumer surplus in each of the simulated scenarios (benchmark, absent Airbnb without price adjustment, absent Airbnb with price adjustment). Then, we compare the consumer surplus in each of the counterfactual scenarios to the benchmark scenario.[39] This comparison allows us to quantify how much consumers gain in terms of consumer surplus in the benchmark scenario with Airbnb compared to the counterfactual scenarios without Airbnb.

[39]The formula for the expected consumer surplus in the nested logit model is an intuitive extension of the well known formula for the logit model and is reported in Appendix A.5.

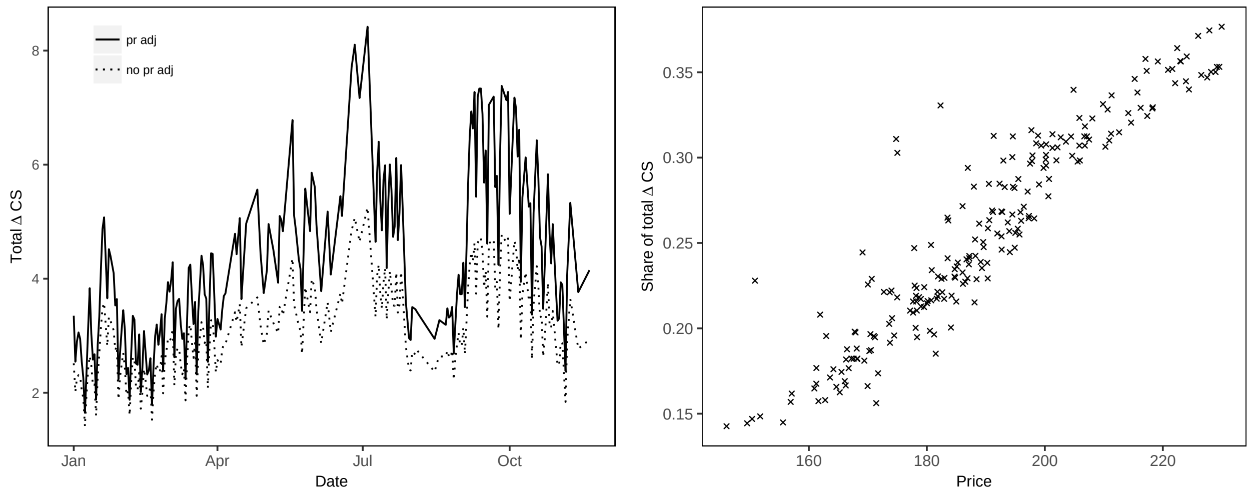
Table 3: Simulated price and occupancy changes for hotels

Star Rating	Price	Δ Pr	Δ Pr %	Δ Q		Δ Q %	
				pr adj	no pr adj	pr adj	no pr adj
1	92.93	4.63	5.00	157.62	144.39	16.28	14.91
2	101.12	10.76	10.64	847.06	944.91	16.08	17.93
3	131.39	18.90	14.39	2269.21	3967.71	10.30	18.01
4	205.92	21.06	10.23	1478.53	3176.55	8.27	17.76
5	470.70	21.04	4.47	169.49	582.18	5.05	17.36

Notes: The second column shows sales-weighted average daily prices when Airbnb is present, the second and third column show resulting price changes when Airbnb is absent. Changes in quantity are calculated for the scenarios with and without price adjustment.

When hotels are allowed to adjust prices, we estimate an average gain in consumer surplus from Airbnb of 31 euro per night. In the counterfactual scenario where hotels are not allowed to adjust prices, we estimate the average gain to be only 23 euro. Therefore, lower hotel prices when Airbnb is present account for approximately 26 percent of the gain in consumer surplus from Airbnb. To estimate the total gain in consumer surplus from Airbnb, we multiply the estimated expected gain in consumer surplus per consumer with the estimated market size. We find that, when we allow hotels to adjust prices absent Airbnb, the average daily gain in consumer surplus from Airbnb is estimated to be approximately 4.3 million euro. Ignoring price adjustments, the gain in total consumer surplus is only 3.1 million euro.

Figure 13a reveals that the gain in consumer surplus from Airbnb is unevenly distributed across markets. The total consumer surplus gain, accounting for hotel price adjustments, ranges from two to eight million euro. In periods of high demand, when hotels charge high prices, the total consumer surplus gain from Airbnb is the highest. Figure 13b shows that the share of the gain in consumer surplus due to hotel price adjustment is higher when hotels charge higher prices in the benchmark scenario. When prices are high even with Airbnb in the market, hotel price responses can account for an aggregate gain in consumer surplus from Airbnb of up to three million euro. Finally, our model also allows us to assess the share of Airbnb consumers that would choose a hotel option absent Airbnb. We find that, accounting for hotel-price adjustments, only 28 percent of Airbnb consumers would book a hotel option if Airbnb were not available. Without price adjustment, this number amounts to 48 percent.



(a) Gain in daily total consumer surplus with and without hotel price adjustment

(b) Share in total consumer surplus gain attributable to hotel price adjustment

Figure 13: Consumer surplus gains from Airbnb with and without hotel price adjustment

6 Conclusion

With the rise of the digital economy, peer-to-peer markets have entered and subsequently disrupted many traditional industries. A prominent example is the rise of Airbnb, from the co-founders hosting their first guests in their apartment in San Francisco in 2007 to featuring more than seven million listings in over 220 countries in 2020.[40] This rapid rise was, however, accompanied by controversial policy debates. One main concern is that Airbnb listings are competing with traditional hotels while subject to much laxer regulation. In this context, it is important to understand the nature of competition between Airbnb and hotels and to

[40] See https://news.airbnb.com/fast-facts/ (accessed: July 1, 2020).

assess how and when consumers profit from the online platform.

In this paper, we first investigate the nature of competition between Airbnb listings and hotels in the city of Paris using data from 2017. We combine granular, daily data on hotel occupancy rates, Airbnb demand, and hotel prices to estimate a two-level nested logit model of demand. Our model allows for taste correlation within the different districts of the city as well as within accommodation types (Airbnb listings or hotels) in each district. Our results suggest that the market is segmented both along geography as well as the type of accommodation. More specifically, we find that cross-price elasticities are larger for accommodations located in the same district. Furthermore, within each district, cross-price elasticities are even larger for accommodations of the same type, i.e. Airbnb listings or hotel rooms.

Based on these parameter estimates, we next ask what the presence of Airbnb implies for consumer surplus. To address this question, we simulate counterfactual scenarios without Airbnb and compare the outcomes with a scenario comparable to the real world. These simulations allow us to compare consumer surplus, hotel revenue, and consumer choices in a world with and without Airbnb. We find that, on average, consumers gain 31 euro per night in terms of consumer surplus from having Airbnb in the market. Taking into account our estimate of the total market size, this gain amounts to a total average gain in consumer surplus of 4.3 million euro per night. This gain in consumer surplus consists of utility gains because of the increased variety due to Airbnb and lower hotel prices due to increased competition. To disentangle the two, we also regard a counterfactual scenario absent Airbnb in which hotels do not adapt prices. Compared to this scenario, average consumer surplus is 23 euro per night higher in the presence of Airbnb. Therefore, the gain from lower prices due to Airbnb amounts to approximately 26 percent of the entire gain in consumer surplus. Consumers profit most in times when demand is high and hotels approach their capacity constraints. In these situations, the aggregate gain in consumer surplus can amount up to eight million euro per night. In these high demand phases, consumers profit most from being able to choose from Airbnb listings because they would otherwise have to choose between hotels asking high prices due to their high occupancy or not booking an accommodation at all. Concerning hotel revenues, our simulations suggest that they would increase by 1.8 million euro per night absent Airbnb.

Our results are in line with those of Farronato and Fradkin (2018) and Zervas et al. (2017). Farronato and Fradkin (2018) find that competition between Airbnb and hotels in the US is particularly fierce when demand is high and hotels are close to capacity-constrained. They document that consumers profit most from Airbnb in these times as hotels charge lower prices than they would absent Airbnb. Zervas et al. (2017) find that average hotel revenue decreased in Texas due to the entry of Airbnb. The main mechanism for this result is that hotels charge lower prices due to the increased competition. We complement these results by investigating related aspects on a more fine-grained level. Rather than using data on the city level, we observe and account for variation in demand within the city. Our descriptive results suggest that demand shocks in neighborhoods within the city are an important factor in the demand for short-term accommodation.

All in all, our results suggest that competition between Airbnb and hotels is only moderate most of the times, but Airbnb listings can be an important substitute for consumers in times of high demand. In these times, consumers are able to profit directly from the more flexible supply of Airbnb listings. In general, consumers therefore do seem to profit substantially from the presence of Airbnb in the market.

There is one main caveat to our analysis: We cannot account for geographic proximity of different districts. Therefore, the relationship between products in different districts is identical in the model, irrespective of their geographic proximity. Methods that allow to model geographic relationships more precisely are either much more computationally involved or introduce additional complexity when handling endogeneity (Fox et al., 2016). Therefore, our decision to use the two-level nested logit is motivated by its tractability, as well as ability to incorporate instrumental variables using standard techniques.

Of course, our paper only analyzes one part of the greater scheme when it comes to regulating Airbnb. In particular, because we do not estimate hotel profits, we cannot say anything about how Airbnb impacts those. Furthermore, our analysis abstracts from externalities that Airbnb can create in other markets such as the housing market (Horn and Merante, 2017; Koster et al., 2018; Garcia-López et al., 2019; Barron et al., 2020; Duso et al., 2020) or on local neighborhoods (Basuroy et al., 2020). Therefore, while our results suggest that Airbnb increases consumer welfare for travelers, it is not clear whether this increase is sufficient to offset other potential welfare losses that are relevant when reflecting on total welfare.

References

Ackerberg, D. A. and Rysman, M. (2005). Unobserved Product Differentiation in Discrete-Choice Models: Estimating Price Elasticities and Welfare Effects. *The RAND Journal of Economics*, 36(4):771–788.

Aguiar, L. and Waldfogel, J. (2018). As streaming reaches flood stage, does it stimulate or depress music sales? *International Journal of Industrial Organization*, 57:278–307.

Barron, K., Kung, E., and Proserpio, D. (2020). The Effect of Home-Sharing on House Prices and Rents: Evidence from Airbnb. *Marketing Science*, forthcoming.

Basuroy, S., Kim, Y., and Proserpio, D. (2020). Estimating the impact of Airbnb on the local economy: Evidence from the restaurant industry. *Working Paper*.

Berry, S., Levinsohn, J., and Pakes, A. (1995). Automobile Prices in Market Equilibrium. *Econometrica*, 63(4):841–890.

Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, 25(4):242–262.

Cachon, G. P., Daniels, K. M., and Lobel, R. (2017). The Role of Surge Pricing on a Service Platform with Self-Scheduling Capacity. *Manufacturing & Service Operations Management*, 19(3):368–384.

Castillo, J. C. (2019). Who Benefits from Surge Pricing? *Working Paper*.

Cohen, P., Hahn, R., Hall, J., Levitt, S., and Metcalfe, R. (2016). Using Big Data to Estimate Consumer Surplus: The Case of Uber. *NBER Working Paper Series*, 22627.

Cramer, J. and Krueger, A. B. (2016). Disruptive Change in the Taxi Business: The Case of Uber. *American Economic Review: Papers & Proceedings*, 106(5):177–182.

de Palma, A., Picard, N., and Waddell, P. (2007). Discrete choice models with capacity constraints: An empirical analysis of the housing market of the greater Paris region. *Journal of Urban Economics*, 62(2):204–230.

Duso, T., Michelsen, C., Schäfer, M., and Tran, K. D. (2020). Airbnb and Rents: Evidence from Berlin. *Mimeo*.

Farronato, C. and Fradkin, A. (2018). The Welfare Effects of Peer Entry in the Accommodation Market: The Case of Airbnb. *NBER Working Paper*, 24361.

Fox, J. T., il Kim, K., and Yang, C. (2016). A simple nonparametric approach to estimating the distribution of random coefficients in structural models. *Journal of Econometrics*, 195(2):236–254.

Garcia-López, M.-À., Jofre-Monseny, J., Mazza, R. M., and Segú, M. (2019). Do Short-term Rental Platforms Affect Housing Markets? Evidence from Airbnb in Barcelona. *IEB*

Working Paper 2019/05.

Guda, H. and Subramanian, U. (2019). Your Uber Is Arriving: Managing On-Demand Workers Through Surge Pricing, Forecast Communication, and Worker Incentives. *Management Science*, 65(5):1995–2014.

Hall, J. D., Palsson, C., and Price, J. (2018). Is Uber a substitute or complement for public transit? *Journal of Urban Economics*, 108:36–50.

Haywood, J., Mayock, P., Freitag, J., Owoo, K. A., and Fiorilla, B. (2017). Airbnb & Hotel Performance. An analysis of proprietary data in 13 global markets. Technical report, STR.

Horn, K. and Merante, M. (2017). Is home sharing driving up rents? Evidence from Airbnb in Boston. *Journal of Housing Economics*, 38:14–24.

Hunold, M., Kesler, R., and Laitenberger, U. (2020). Rankings of Online Travel Agents, Channel Pricing, and Consumer Protection. *Marketing Science*, 39(1):92–116.

Koster, H., van Ommeren, J., and Volkhausen, N. (2018). Short-term rentals and the housing market: Quasi-experimental evidence from Airbnb in Los Angeles. *CEPR Discussion Paper*, DP13094.

Kroft, K. and Pope, D. G. (2014). Does Online Search Crowd Out Traditional Search and Improve Matching Efficiency? Evidence from Craigslist. *Journal of Labor Economics*, 32(2):259–303.

Lam, C. T., Liu, M., and Hui, X. (2020). The Geography of Ridesharing: A Case Study of New York City. *Working Paper*.

McFadden, D., Tye, W. B., and Train, K. (1977). An Application of Diagnostic Tests for the Independence From Irrelevant Alternatives Property of the Multinomial Logit Model. *Transportation Research Record*, 637:39–46.

Neeser, D. (2015). Does Airbnb Hurt Hotel Business: Evidence from the Nordic Countries. *Mimeo*.

Seamans, R. and Zhu, F. (2014). Responses to entry in multi-sided markets: The impact of Craigslist on local newspapers. *Management Science*, 60(2):476–493.

Valentin, M. (2020). Regulating Short-Term Rental Housing: Evidence from New Orleans. *Real Estate Economics*, forthcoming.

Verboven, F. (1996). International price discrimination in the European car market. *The RAND Journal of Economics*, 27(2):240–268.

Zervas, G., Proserpio, D., and Byers, J. W. (2017). The Rise of the Sharing Economy: Estimating the Impact of Airbnb on the Hotel Industry. *Journal of Marketing Research*, 54(5):687–705.

A Appendix

A.1 Descriptives

Table 4: Prices, occupancy, and offer

Quality	Price		Rooms Occupied		Rooms Offered	
	Hotels	Airbnb	Hotels	Airbnb	Hotels	Airbnb
1	81.33	41.43	1112.61	4214.60	1583.48	12577.39
	(15.73)	(1.60)	(186.98)	(847.23)	(11.37)	(1267.82)
2	99.63	64.03	6599.36	4258.33	8895.86	12885.61
	(17.00)	(1.94)	(1228.55)	(849.14)	(86.78)	(1304.51)
3	125.93	92.67	24714.45	4550.64	31748.35	14160.72
	(23.02)	(2.92)	(3906.06)	(964.24)	(131.52)	(1291.13)
4	192.81	172.83	19708.64	4539.22	25502.31	15036.03
	(31.89)	(8.07)	(3093.73)	(1138.67)	(388.45)	(1118.71)
5	485.10	-	4579.63	-	6483.31	-
	(79.59)	-	(796.29)	-	(225.32)	-

Notes: The table shows averages with standard deviations in parentheses. The prices shown are average prices per night. For hotels, prices are calculated based on dates for which less than 65 percent of prices are missing.

Table 5: Matched vs. non-matched hotels: Hotel level

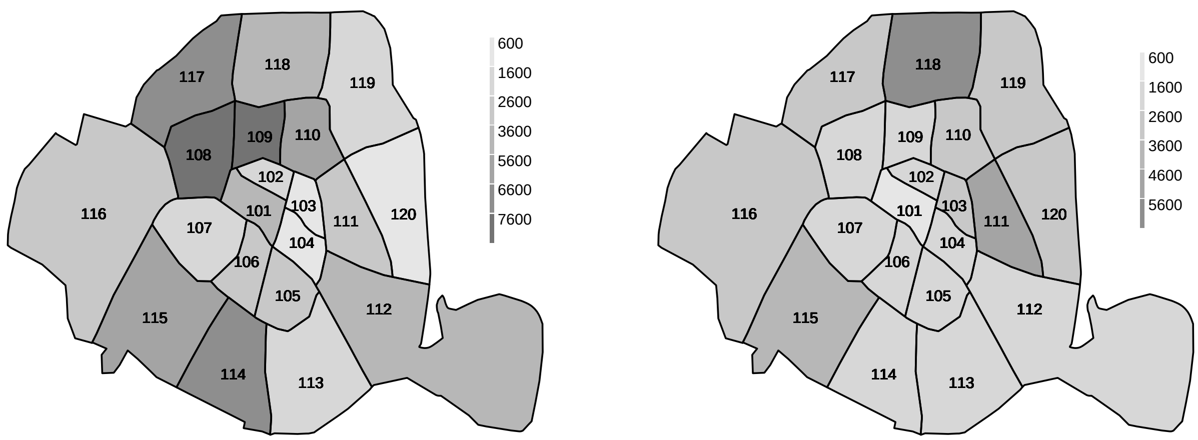
Star Rating		Rooms Occupied		Rooms Offered	
	Share	Matched	Non-matched	Matched	Non-matched
1	33.33	45.39	21.08	63.93	31.26
		(77.89)	(10.52)	(102.47)	(14.53)
2	26.75	28.69	31.39	38.77	41.73
		(33.86)	(33.31)	(43.36)	(41.80)
3	32.26	33.85	36.75	43.37	47.11
		(32.44)	(39.26)	(40.16)	(52.17)
4	45.13	59.91	51.63	78.10	65.75
		(96.16)	(68.56)	(125.52)	(83.51)
5	32.89	58.18	74.35	82.95	103.65
		(39.51)	(66.74)	(53.78)	(95.2)

Notes: The second column shows the percentage of hotels that we could not match for each category. The other statistics are the average daily number of occupied and offered rooms per hotel, calculated over the entire year (standard deviations in parentheses). The unit of observation is a single hotel.

Table 6: Matched vs. non-matched hotels: Room level

Star Rating		Rooms Occupied		Rooms Offered	
	Share	Matched	Non-matched	Matched	Non-matched
1	19.25	907.73	205.50	1278.78	304.71
		(153.63)	(34.94)	(8.94)	(9.13)
2	27.79	4753.02	1859.66	6423.65	2472.21
		(895.84)	(350.75)	(49.79)	(46.05)
3	35.84	15900.61	8877.17	20370.08	11378.27
		(2579.40)	(1397.01)	(58.68)	(131.76)
4	40.25	11686.96	8061.20	15236.77	10265.54
		(1881.30)	(1305.69)	(522.86)	(387.19)
5	35.48	2934.05	1650.19	4182.78	2300.53
		(488.89)	(321.76)	(60.70)	(171.71)

Notes: This table is similar to Table 5. However, rather than comparing hotel-level averages, we compare aggregate room capacities and occupation by hotel category. The second columns shows the percentage of non-matched rooms for each hotel category. The other statistics are the average daily number of occupied and offered rooms per hotel category, calculated over the entire year (standard deviations in parentheses). This table shows that we capture the majority of hotel rooms in the sample of matched hotels.



(a) Hotels (b) Airbnb

Figure 14: Average geographical distribution of hotel rooms and Airbnb listings over the year. The legends show the lower limit of brackets, i.e. “600” represents districts with Airbnb or hotel numbers between 600 and 1,600 rooms. The scales in both figures were chosen to allow for comparison between both accommodation types. The minimum hotel capacity in a district is 608, the maximum 8524. The minimum Airbnb offer in a district is 1338, the maximum 6130. In specification (5) of Table 1, we include hotel-specific fixed effects for the following districts: 101, 108, 109, 114, 115 and 117.

A.2 Market Size as Instrument

In this appendix, we show by simulation that market size can act as a valid instrument. We have four single-product firms. For each firm j, the econometrician observes two characteristics x_j and p_{jt}. In each market t, we have N_t individuals that either choose one of the four products or the outside good $j = 0$ with utility normalized to zero. Note that the market size varies by market. For 500 markets, we simulate the following scenario:

1 We set the vector of products characteristics x equal to $(0.4, 0.5, 0.9, 1)$ and the price vector p equals to $(0.2, 0.3, 0.4, 0.5)$.

2 Utility of consumers is given by $u_{ij} = x_j - 2 \times p_{jt} + \mathcal{E}_{jt} + \epsilon_{jt}$, where $\mathcal{E}_{jt}$ are unobserved product- and market-specific shocks and ϵ_{jt} is i.i.d extreme value type 1.

3 For each market t, the market size N_t is randomly drawn from the vector (500, 600, 700, 800, 1000).

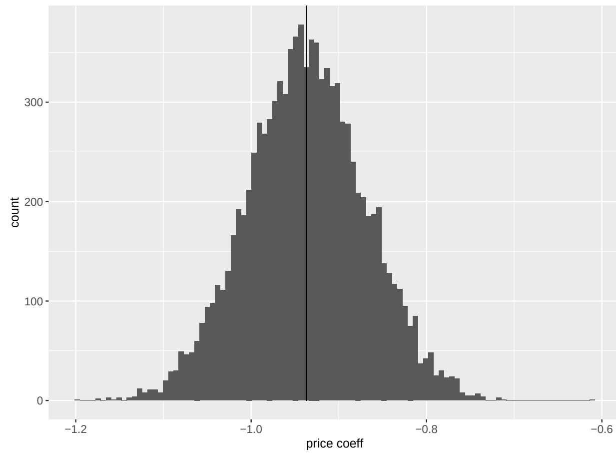
4 Firms observe the market size N_t and set prices according to $p_{jt} = p_j + 0.001 \times N_t$

5 Additionally in each period *one* product j is randomly shocked with $\mathcal{E}_{jt} \sim U[0, 1]$. If product j is shocked, the firm sets the price $p_{jt} = p_j + 0.001 \times N_t + 0.1 \times \mathcal{E}_{jt}$

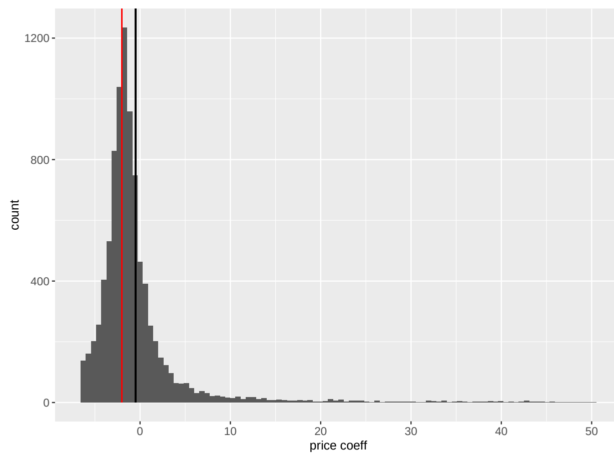
The above situation emulates a scenario where firms respond to changes in the market size by uniformly raising prices. Additionally, in each period, one firm experiences a demand shock that it observes and raises prices accordingly. This emulates the idea of localized demand shocks. We simulate the aggregate market outcomes and perform three different estimations: (a) a simple logit, (b) an IV-logit with classical BLP instruments, i.e. the mean over rivals' x_{jt}, and (c) an IV-logit with the BLP instruments interacted with the market size N_t.

Note that the classical BLP instrument is collinear with the vector x in this setting. The reason is that the classical BLP instruments only work when either the choice set or the characteristics of the products vary by market. Therefore, in our simulation, we randomly drop one product in each market situation. When interacting the BLP instrument with the market size, this collinearity does not arise.

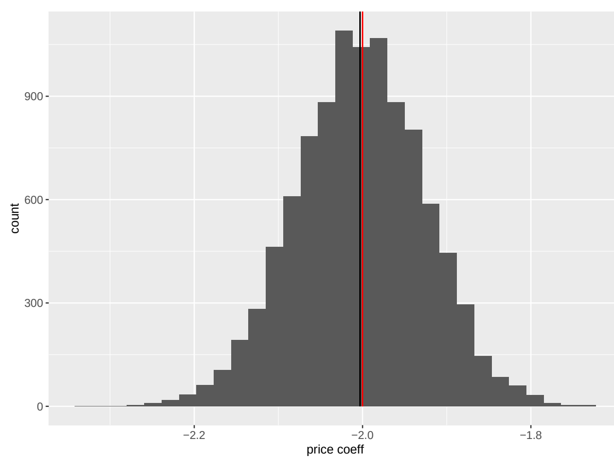
We perform 10,000 Monte Carlo runs. In each run, we simulate the aggregate choices and estimate the structural parameters. Figure 15a shows the histogram for the estimated price coefficient in each of the simulation runs for the logit model. Recall that the true price coefficient is -2. As expected, in the naive model, we underestimate the magnitude of the true price parameter. The bias is approximately equal to 1.4, on average. Figure 15b shows the result for the classical BLP instrumental variable technique. The performance of the estimator is surprisingly poor. In some simulation runs, we obtain extremely large deviations from the true parameter value. Figure 15b therefore only shows the distribution of values while truncating outliers. The mean over the Monte Carlo runs is equal to 3.99 with a variance of 218,443. The median is equal to -1.81. By contrast, the estimator that interacts the classical BLP instruments with the market size is well-behaved and appears to be consistent with a mean of -2.003, a variance of 0.006 and a median of -2.002 (see Figure 15c).



(a) Logit



(b) IV-logit “BLP”



(c) IV-logit “BLP $\times$ N_t”

Figure 15: Histograms of estimated price coefficients. The red line shows the true price coefficient, the black line shows the mean estimated price coefficients.

A.3 First-stage F-statistics

Table 7 reports the F-statistics for the excluded instruments for the various specifications reported in Table 1. We report the F-statistic obtained for each endogenous variable in each specification. Note that the F-statistic for price in the specification of column (2) of Table 1 is identical to the F-statistic for the price in the specification of column (3) of Table 1.

Table 7: F-statistics corresponding to Table 1

	(2)/(3) Logit/NL	(4) NL	(5) NL
Price	602.83	648.81	3631.20
σ_1	257.26	263.96	4906.98
σ_2	7636.45	8045.26	11793.30

A.4 Robustness Checks

Table 8: Results with alternative market size definition

	(1) Logit	(2) Logit (IV)	(3) NL	(4) NL	(5) NL
Price	0.002***	−0.008***	−0.006***	−0.007***	−0.005***
	(0.0001)	(0.0004)	(0.0003)	(0.0003)	(0.0003)
σ_1			0.770***	0.645***	0.712***
			(0.018)	(0.014)	(0.018)
σ_2			0.234***	0.228***	0.207***
			(0.010)	(0.009)	(0.006)
Airbnb category 2	0.289***	0.506***	0.065***	0.140***	−0.015
	(0.041)	(0.045)	(0.022)	(0.023)	(0.011)
Airbnb category 3	0.420***	0.935***	0.278***	0.394***	0.084***
	(0.040)	(0.047)	(0.028)	(0.028)	(0.018)
Airbnb category 4	0.259***	1.617***	0.782***	0.942***	0.469***
	(0.042)	(0.068)	(0.045)	(0.045)	(0.039)
1-star hotel	−0.544***	−0.144***	0.742***	0.602***	2.576***
	(0.048)	(0.053)	(0.035)	(0.031)	(0.066)
2-star hotel	0.739***	1.333***	1.000***	1.037***	2.744***
	(0.041)	(0.049)	(0.025)	(0.026)	(0.094)
3-star hotel	2.065***	2.942***	1.483***	1.699***	3.275***
	(0.041)	(0.055)	(0.042)	(0.041)	(0.121)
4-star hotel	1.641***	3.217***	1.854***	2.073***	3.553***
	(0.043)	(0.076)	(0.052)	(0.051)	(0.127)
5-star hotel	−0.220***	4.427***	3.287***	3.526***	4.537***
	(0.067)	(0.180)	(0.117)	(0.113)	(0.167)
Log Airbnb offer					0.356***
					(0.016)
Constant	−6.486***	−6.036***	−4.174***	−4.842***	−6.866***
	(0.039)	(0.044)	(0.046)	(0.071)	(0.128)
Market FEs	*N*	*N*	*N*	*Y*	*Y*
N	33,504	33,504	33,504	33,504	33,504
Adjusted R^2	0.374	0.120	0.641	0.697	0.861

Notes: These results correspond to the main analysis but using the instantaneous market size measure obtained from the Google Search data. The lowest Airbnb quality category serves as the base category. *, **, *** indicate statistical significance at the ten, five, and one percent level, respectively.

Table 9: Results using period from February 1st to July 1st

	(1)	(2)	(3)	(4)	(5)
	Logit	Logit (IV)	NL	NL	NL
Price	0.002***	−0.009***	−0.007***	−0.007***	−0.005***
	(0.0002)	(0.001)	(0.0005)	(0.0004)	(0.0004)
σ_1			0.703***	0.666***	0.786***
			(0.026)	(0.020)	(0.027)
σ_2			0.224***	0.220***	0.211***
			(0.016)	(0.013)	(0.009)
Airbnb category 2	0.317***	0.556***	0.128***	0.147***	−0.025
	(0.062)	(0.069)	(0.036)	(0.034)	(0.018)
Airbnb category 3	0.472***	1.033***	0.387***	0.412***	0.081***
	(0.061)	(0.073)	(0.046)	(0.042)	(0.028)
Airbnb category 4	0.311***	1.783***	0.966***	0.985***	0.474***
	(0.064)	(0.107)	(0.074)	(0.066)	(0.055)
1-star hotel	−0.513***	−0.079	0.725***	0.716***	2.369***
	(0.072)	(0.080)	(0.055)	(0.045)	(0.101)
2-star hotel	0.907***	1.553***	1.158***	1.167***	2.444***
	(0.062)	(0.075)	(0.042)	(0.039)	(0.145)
3-star hotel	2.213***	3.169***	1.755***	1.803***	2.870***
	(0.062)	(0.086)	(0.069)	(0.061)	(0.185)
4-star hotel	1.778***	3.464***	2.146***	2.181***	3.156***
	(0.065)	(0.120)	(0.085)	(0.075)	(0.191)
5-star hotel	−0.104	4.782***	3.674***	3.657***	4.145***
	(0.098)	(0.278)	(0.185)	(0.163)	(0.237)
Log Airbnb offer					0.288***
					(0.024)
Constant	−6.667***	−6.179***	−4.423***	−4.685***	−6.184***
	(0.060)	(0.069)	(0.074)	(0.087)	(0.191)
Market FEs	*N*	*N*	*N*	*Y*	*Y*
N	14,773	14,773	14,773	14,773	14,773
Adjusted R^2	0.401	0.160	0.638	0.721	0.879

Notes: These results correpond to the main analysis, but using only the sample from February to July 2017. Figure 1 shows that during this period, the number of Airbnb listings varies less than in the second half of the year. The lowest Airbnb quality category serves as the base category. *, **, *** indicate statistical significance at the ten, five, and one percent level, respectively.

A.5 Formulas for Elasticities and Consumer Surplus

In this section, we present the formulas for the own- and cross-price elasticities in the two-level nested logit model and the consumer surplus in logit and nested logit models.

A.5.1 Elasticities

We start with the formulas of the cross-price elasticities. We follow the notation used by Verboven (1996):

$$-e_{jj} = \frac{\partial q_j}{\partial p_j}\frac{p_j}{q_j} = -\alpha p_j\Big(\frac{1}{1-\sigma_1} - (\frac{1}{1-\sigma_1} - \frac{1}{1-\sigma_2})\frac{s_j}{s_{hd}} - \frac{\sigma_2}{1-\sigma_2}\frac{s_j}{s_d} - s_j\Big) \tag{7}$$

$$e_{jk} = \frac{\partial q_k}{\partial p_j}\frac{p_j}{q_k} = \alpha p_j\Big((\frac{1}{1-\sigma_1} - \frac{1}{1-\sigma_2})\frac{s_j}{s_{hd}} + \frac{\sigma_2}{1-\sigma_2}\frac{s_j}{s_d} + s_j\Big) \tag{8}$$

$$e_{jk'} = \frac{\partial q_{k'}}{\partial p_j}\frac{p_j}{q_{k'}} = \alpha p_j\Big(\frac{\sigma_2}{1-\sigma_2}\frac{s_j}{s_d} + s_j\Big) \tag{9}$$

$$e_{jk''} = \frac{\partial q_{k''}}{\partial p_j}\frac{p_j}{q_{k''}} = \alpha p_j s_j \tag{10}$$

A.5.2 Consumer Surplus

In a two-level nested logit model, the formula for the expected utility of consumers is given by:

$$CS = ln\Big(\sum_{d=0}^{D} e^{I_{dt}}\Big) \tag{11}$$

Where I_{dt} is defined by:

$$I_{dt} = (1-\sigma_2)ln\Big(\sum_{h \in V_{dt}} e^{I_{hdt}/(1-\sigma_2)}\Big) \tag{12}$$

and I_{hdt} is defined by:

$$I_{hdt} = (1-\sigma_1)ln\Big(\sum_{l \in V_{hdt}} e^{\delta_{jt}/(1-\sigma_1)}\Big) \tag{13}$$

If we multiply Equation (11) by the negative inverse of the estimated price coefficient, we obtain the euro value of the consumer surplus. The formula for the consumer surplus is a natural extension of the consumer surplus for the simple logit. Equation (11) is the logsum over the denominators of the nests (which are multiplied with the nest parameters). Each denominator of a nest is itself a logsum over the denominators of the subnests contained in the nest (multiplied with the subnest parameters).

A.6 Semi-Elasticities

Table 10: Average estimated demand semi-elasticities $\times 100$

			Cross-price elasticities		
Type	Category	Own-price elasticity	Same district & Type	Same district	Other district
Airbnb	1	-1.3217	0.4591	0.0161	0.0013
	2	-1.5037	0.2771	0.0093	0.0007
	3	-1.4810	0.2998	0.0098	0.0008
	4	-1.4058	0.3750	0.0111	0.0009
Hotels	1	-1.6924	0.0884	0.0071	0.0007
	2	-1.6476	0.1332	0.0110	0.0011
	3	-1.2696	0.5112	0.0431	0.0042
	4	-1.3463	0.4345	0.0360	0.0033
	5	-1.3986	0.3822	0.0353	0.0039